

**CHARACTERIZATION OF PIGMENT PRODUCTION BY  
*ESCHERICHIA COLI* CONTAINING A CLONED  
*RHODOCOCCUS* GENE**

by

**STEPHEN LEWIS HART**

A dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in the Faculty of Science, University of Cape Town.

May 1991

The University of Cape Town has been given  
the right to reproduce this thesis in whole  
or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

## **CERTIFICATION OF SUPERVISOR**

In terms of paragraph GP 8 of "General Regulations for the degree of Doctor of Philosophy (PhD)" I, as supervisor of the candidate S. L. Hart, certify that I approve of the incorporation in this thesis of material that has already been published or submitted for publication.

**Signature Removed**

**Professor D. R. Woods  
Deputy Vice Chancellor (Research)  
Director of Microbiology  
Research Units**

## ACKNOWLEDGEMENTS

I wish to thank Professor David Woods for his attentive supervision during the course of these studies.

My thanks go to Professor Doug Rawlings, not only for helpful scientific discussions but also for many enjoyable games of squash.

Some of the nucleotide sequencing was done by Di James and her efficient, friendly assistance was much appreciated.

I am pleased to thank my very good friends and colleagues in the department for creating a working atmosphere which was both stimulating and laced with much good humour.

The NMR spectroscopy was performed by Professor Klaus Koch of the Department of Chemistry, University of Cape Town. I am especially grateful to him for his assistance in analysing the structure of the pigments.

This work would not have been possible without the financial assistance of the Foundation for Research and Development.

**TABLE OF CONTENTS**

|                         |   |            |
|-------------------------|---|------------|
|                         | <b>ABSTRACT</b>   | <b>II</b>  |
|                         | <b>ABBREVIATIONS</b>  | <b>IV</b>  |
| <b>CHAPTER ONE</b>      | <b>GENERAL INTRODUCTION</b>   | <b>1</b>   |
| <b>CHAPTER TWO</b>      | <b>ANALYSIS OF THE PIGMENTS AND THE<br/>PATHWAY OF PIGMENT PRODUCTION</b> | <b>21</b>  |
| <b>CHAPTER THREE</b>    | <b>NUCLEOTIDE SEQUENCE OF THE<br/>PIGMENT GENE</b>                        | <b>44</b>  |
| <b>CHAPTER FOUR</b>     | <b>CONSTRUCTION OF AN INSERTIONAL-<br/>INACTIVATION CLONING VECTOR</b>    | <b>65</b>  |
| <b>APPENDIX A</b>       | <b>STANDARD METHODS</b>   | <b>88</b>  |
| <b>APPENDIX B</b>       | <b>BUFFERS AND MEDIA</b>  | <b>93</b>  |
| <b>APPENDIX C</b>       | <b>PRELIMINARY NMR PIGMENT STUDY</b>                                      | <b>97</b>  |
| <b>APPENDIX D</b>       | <b>LIST OF VECTORS</b>  | <b>109</b> |
| <b>LITERATURE CITED</b> |   | <b>111</b> |

CHARACTERIZATION OF PIGMENT PRODUCTION BY  
*ESCHERICHIA COLI* CONTAINING A CLONED *RHODOCOCCUS*  
GENE

by

Stephen Lewis Hart

Department of Microbiology, University of Cape Town,  
Rondebosch, 7700, South Africa

May, 1991

ABSTRACT

The screening of a *Rhodococcus* genebank in *Escherichia coli* lead to the discovery of a pigment-producing clone. Initial studies revealed a blue component and a pink component in bacterial pigment preparations and suggested that they were different from the prodigiosins and actinorhodins produced by other actinomycetes. This dissertation represents a further analysis of the genetics and biochemistry of pigment production, and the development of an insertional-inactivation cloning vector utilizing the *Rhodococcus* pigment gene.

The DNA of the pigment-producing clone was analysed by restriction mapping and sequenced. A single *Rhodococcus* gene of 1.1 kbp was found to be responsible for pigment production in *E. coli*. This gene had a putative ribosome binding site and coded for an enzyme of  $M_r$  42,560. Deletion analysis and *in vitro* transcription-translation experiments supported the hypothesis that pigment production in *E. coli* was due to a single enzyme. The gene and gene product did not show any similarity when compared with nucleotide and protein sequences in computer data bases.

Evidence was obtained from bacterial studies that the pigment pathway involved the conversion of tryptophan to indole by tryptophanase of *E. coli* and then to the pigment by the action of the cloned *Rhodococcus* gene. The solubility properties of the pigment, TLC analysis and NMR and visible spectrophotometry supported the hypothesis that the blue pigment was indigo and that the red pigment was indirubin, an isomer of indigo.

An insertional-inactivation cloning vector, pSLH8, was developed containing the pigment gene as a marker. The advantages of this vector were that it produced pigmented colonies on LB agar plates without any further requirement for expensive substrates such as X-Gal, which is required by the pUC series of vectors, and it is capable of being used in any *E. coli* strain whereas the pUC plasmids require special mutant strains for the detection of recombinants.

## ABBREVIATIONS

|                   |                                     |
|-------------------|-------------------------------------|
| A                 | Adenine                             |
| Ap                | ampicillin                          |
| ATCC              | American Type Culture Collection    |
| <i>bla</i>        | $\beta$ -lactamase                  |
| Bq                | Becquerel                           |
| bp                | base pair(s)                        |
| C                 | cytosine                            |
| <i>cat</i>        | chloramphenicol acetyl transferase  |
| Ci                | Curie                               |
| Cm                | chloramphenicol                     |
| CsCl              | caesium chloride                    |
| DMSO              | dimethyl sulphoxide                 |
| DNA               | deoxyribonucleic acid               |
| dNTP              | deoxynucleotide triphosphate        |
| EDTA              | ethylenediaminetetra-acetic acid    |
| EtdBr             | ethidium bromide                    |
| <i>g</i>          | standard gravitational acceleration |
| G                 | guanine                             |
| G+C               | guanine and cytosine                |
| h                 | hour(s)                             |
| kb                | kilobase(s)                         |
| kbp               | kilobase pair(s)                    |
| LB                | Luria-Bertani                       |
| min               | minutes                             |
| $M_r$             | relative molecular mass             |
| n                 | nanno                               |
| NMR               | nuclear magnetic resonance          |
| OD <sub>600</sub> | optical density at 600 nm           |
| ORF               | open reading frame                  |
| p                 | plasmid                             |
| $P_R$             | lambda rightward promoter           |
| PAGE              | polyacrylamide gel electrophoresis  |
| pv.               | pathovar                            |



|            |                                   |
|------------|-----------------------------------|
| r          | (superscript) resistance          |
| $R_F$      | relative migration rate by TLC    |
| RNA        | ribonucleic acid                  |
| RNase      | ribonuclease                      |
|            |                                   |
| s          | second(s)                         |
| s          | (superscript) sensitivity         |
| SDS        | sodium dodecyl sulphate           |
| sp.        | species                           |
| spp.       | species (plural)                  |
|            |                                   |
| T          | thymine                           |
| Tc         | tetracycline                      |
| TLC        | thin layer chromatography         |
| Tris       | tris (hydroxymethyl) aminomethane |
|            |                                   |
| U          | units of enzyme activity          |
| UV         | ultra violet                      |
|            |                                   |
| v/v        | volume/volume                     |
|            |                                   |
| w/v        | weight/volume                     |
|            |                                   |
| $\alpha$   | alpha                             |
| $\beta$    | beta                              |
| $\Delta$   | delta                             |
| $\epsilon$ | epsilon                           |
| $\phi$     | phi                               |
| $\lambda$  | lambda                            |
| $\mu$      | micro                             |

## CHAPTER ONE

### GENERAL INTRODUCTION

|  |    |
|--|----|
| 1.1 THE PROPERTIES AND USES OF <i>RHODOCOCCUS</i>  | 2  |
| 1.1.1 Characteristics  | 2  |
| 1.1.2 Metabolism: biotransformations   | 2  |
| 1.1.3 Metabolism: degradation  | 3  |
| 1.2 THE GENETICS OF <i>RHODOCOCCUS</i>   | 5  |
| 1.2.1 Genetic recombination  | 5  |
| 1.2.2 Phages of <i>Rhodococcus</i>   | 6  |
| 1.2.3 Naturally occurring plasmids   | 7  |
| 1.2.4 New techniques and plasmids  | 9  |
| 1.2.5 Gene cloning studies   | 10 |
| 1.3 BACTERIAL PIGMENTS   | 10 |
| 1.3.1 Pigments of <i>Rhodococcus</i>   | 11 |
| 1.3.2 Pigments of <i>Streptomyces</i>  | 11 |
| 1.3.3 Pigment production by other bacteria   | 12 |
| 1.3.4 Pigment production by <i>E. coli</i>   | 13 |
| 1.4 THE PRODUCTION OF INDIGO   | 14 |
| 1.4.1 Historical   | 14 |
| 1.4.2 Preparation and synthesis  | 15 |
| 1.4.3 Bacteria and indigo  | 16 |
| 1.4.4 Aromatic dioxygenases and indigo production from recombinant <i>E. coli</i> clones | 17 |

## CHAPTER ONE

### GENERAL INTRODUCTION

#### 1.1 THE PROPERTIES AND USES OF *RHODOCOCCUS*

##### 1.1.1 Characteristics

The genus *Rhodococcus* was described in detail after a numerical taxonomic classification study assigned generic status to a group of identifiable bacterial strains previously assigned to the genera *Gordona*, *Mycobacterium* and *Nocardia* (Goodfellow and Alderson, 1977). Nocardioform bacteria of the genus *Rhodococcus* comprise aerobic, non-motile, Gram-positive actinomycetes that show considerable morphological diversity and may be partially acid-fast cocci or short rods. Growth occurs by the formation of a primary mycelium that soon fragments into irregular elements. Some strains form feeble aerial hyphae but they do not form endospores or conidia. Colonies may be rough, smooth, mucoid or mycobacteria-like and are usually pigmented buff, orange, pink or red, although colourless variants do occur. Most strains grow optimally at 30 °C and some require thiamine for growth. The cell wall peptidoglycan contains *meso*-diaminopimelic acid, arabinose and galactose. The G+C content of the DNA ranges from 59-69 mol %.

Species have a wide distribution occurring in the soil, in the intestinal flora of insects, in cow dung, manured meadows and fresh water habitats and are often isolated from enrichments with both aliphatic and aromatic hydrocarbons. The type species is *Rhodococcus rhodochrous* (Goodfellow and Alderson, 1977).

##### 1.1.2 Metabolism: biotransformations

Bacteria of the genus *Rhodococcus* display great metabolic diversity comparable to that of the Gram-negative pseudomonads (Cross, 1982). Most notably, they are capable of a wide range of biotransformations, i.e. the production of novel derivatives from known substrates. Some of the chemical modifications performed are highly selective involving complex substrates. Numerous examples of

biotransformations are described in reviews (Peczynska-Czoch and Mordarski, 1984; Tarnok, 1976). *Rhodococcus* spp. are capable of converting cholesterol to substances which are precursors of steroid hormones or oral contraceptives (Ferreira *et al.*, 1984). A strain of *Rhodococcus* which produces high levels of L-phenylalanine hydrogenase was isolated and the enzyme was purified and used in a process for the production of L-phenylalanine (Hummel *et al.*, 1987). Another *Rhodococcus* sp. was isolated which used squalene as a sole carbon source producing a novel unsaturated ketone (Setchell *et al.*, 1985). The optimized production of 1,2-epoxytetradecane from 1-tetradecane was reported (Furuhashi *et al.*, 1981; Furuhashi and Takagi, 1984). *Rhodococcus* N-774 is capable of converting acrylonitrile into acrylamide by virtue of its high nitrile hydratase activity and is being developed for industrial production of acrylamide (Asano *et al.*, 1982; Watanabe *et al.*, 1987a; 1987b). Yet another *Rhodococcus* strain was capable of degrading acrylamide which can be a highly toxic waste (Arai *et al.*, 1981). A strain of *R. rhodochrous* which has high nitrilase activity was shown to be capable of the conversion of 3-cyanopyridine to nicotinic acid, a vitamin which is used in medicine and as an animal feed supplement (Mathew *et al.*, 1988). Another strain of *R. rhodochrous* produces a novel nitrilase which degrades aliphatic nitriles (Kobayashi *et al.*, 1990). Nitrile compounds are synthesized on a large scale as solvents, plastics, sythetic rubber, pharmaceuticals, herbicides and starting materials for other industrially important chemicals. Unfortunately, they are also extremely toxic, not only to the central nervous system but as mutagens. They are now recognised as a serious form of environmental pollution. The nitrilase-producing *R. rhodochrous* in a stable activated sludge system would constitute a major improvement on the treatment of industrial wastes containing aliphatic nitriles where the conventional, activated sludge systems are highly susceptible to inactivation by the toxic wastes (Kobayashi *et al.*, 1990).

### 1.1.3 Metabolism: degradation

*Rhodococcus* spp. have been shown to degrade many toxic, organic industrial waste products as well as pesticide and herbicide residues which would otherwise persist in the environment. Pollutants such as oil, biocides and carcinogenic hydrocarbons

may be subjected to microbial attack by a range of microorganisms including the rhodococci, converting the toxins to harmless products.

Chlorinated guaiacols and chlorinated phenols are highly toxic waste products of the wood preserving industry. Chlorinated guaiacols are formed during chlorine bleaching of wood pulp and released into the environment in used water where they accumulate. Chloroguaicol concentrations of 1 to 100 g l<sup>-1</sup> have been found in lakes and 1 to 2 mg kg<sup>-1</sup> dry weight in lake sediments. In addition, chloroguaiacols are formed by the combustion of organic matter and consequently have been found in rain water. *In vitro* studies have shown that chloroguaiacols uncouple oxidative phosphorylation and disturb conjugate formation by rat liver microsomes. It is thus important to know the fate of these compounds in the environment. A strain of *Rhodococcus chlorophenolicus* was isolated which was capable of degrading chlorinated guaiacols by demethylation and dechlorination (Häggbloom *et al.*, 1986). The same strain was shown to degrade polychlorinated phenols to carbon dioxide (Apajalahti and Salkinoja-Salonen, 1986). Chlorophenol contamination of the environment is extensive. Soil in the vicinity of wood preserving facilities, the major users of chlorophenols, was shown to contain up to a thousand milligrams of chlorophenols per kilogram. Chlorophenols are highly toxic to all forms of life and efficient microbial degradation of these compounds was unknown until the discovery of *R. chlorophenolicus* (Apajalahti and Salkinoja-Salonen, 1986). The degradation pathway of both chloroguaiacols and chlorophenols was initiated by hydroxylation of the benzene ring (Häggbloom *et al.*, 1986; Apajalahti and Salkinoja-Salonen, 1986).

Hydrolytic dechlorination of the aromatic ring of deethylsimazine, a product of the aerobic degradation of the s-triazine herbicides simazine and atrazine, was shown to be carried out by *Rhodococcus corallinus* (Cook and Hütter, 1986).

A strain of *Rhodococcus erythropolis* was isolated which was able to utilize a wide range of halogenated aliphatic compounds as sole sources of carbon and energy (Sallis *et al.*, 1990). These compounds are used widely in agriculture and industry even though many are highly toxic, carcinogenic, teratogenic and persist in the environment. This strain of *Rhodococcus* may be useful for the detoxification of such xenobiotic compounds.

Phthalate esters which are used in industry as plasticisers are also toxic and a problem in industrial wastes. A strain of *R. erythropolis* was isolated which degrades these compounds by hydroxylation (Kurane *et al.*, 1979; 1980). The enzyme which catalyses this reaction was purified and partially characterised (Kurane *et al.*, 1984). Some strains of *Rhodococcus* were reported to degrade acrylamide (Arai *et al.*, 1981), phenolic compounds (Haider *et al.*, 1981) and insecticides such as aldrin (Ferguson *et al.*, 1981). The degradation of the pesticide dalapon (Hirsch and Alexander, 1960) and the conversion of DDT (1,1,1-trichloro-2,2-di-4-chlorophenylethane) to DDD (1,1-dichloro-2,2-di-4-chlorophenylethane) by *R. erythropolis* was reported (Chacko *et al.*, 1966). Warfarin is a Vitamin K antagonist and was used widely as an oral anticoagulant and as a rodenticide. A strain of *Rhodococcus* was shown to specifically and stereoselectively convert *s*-warfarin to the corresponding *s*-alcohol (Davis and Rizzo, 1981).

Strains of *Rhodococcus* are capable of degrading lignin and lignin-related compounds, substrates which very few prokaryotes are capable of utilizing (Crawford *et al.*, 1981; Jaeger *et al.*, 1981; Kuwuhara, 1981; Rast *et al.*, 1980; Trojanowski *et al.*, 1977).

Clearly, the enormous metabolic diversity of the rhodococci is likely to result in these species becoming increasingly important in many areas of biotechnology.

## 1.2 THE GENETICS OF RHODOCOCCLUS

### 1.2.1 Genetic recombination

Probably the first genetic study of *Rhodococcus* species involved the demonstration of genetic recombination between two different species of *Rhodococcus*, formerly known as *Nocardia canicruria* and *N. erythropolis* (Adams, 1964; Adams and Bradley, 1963). Nutritionally complementary mutants isolated from different and identical starting strains of *Rhodococcus* were crossed by simply incubating different combinations of cell suspensions together. The isolation of single prototrophic colonies demonstrated that genetic recombination had taken place. Crosses between different mutant strains of the same species were infertile whereas crosses between different mutant strains of heterologous origin were fertile and prototrophic

recombinants were obtained (Adams, 1963). This observation suggested the existence of a mating compatibility system which governed genetic recombination. Two non-allelic, chromosomal genes, *mat-E* and *mat-C*, appeared to be required for mating and transformation between compatible *Rhodococcus* strains although neither the mechanism of recombination nor the role of the mating factors was understood (Brownell and Denniston, 1984; Brownell and Kelly, 1969). Studies with the nocardiphage  $\phi$ EC, existing as a plasmid, showed that its transfer between *Rhodococcus* spp. was subject to the same *Rhodococcus* compatibility mechanism (Brownell and Adams, 1976). The discovery of genetic recombination between *Rhodococcus* strains led to the construction of a linkage map of more than 65 genetic traits including loci for mating factors and phage integration and resistance sites by the use of protoplast fusions (Adams, 1974; Brownell, 1976; Brownell and Denniston, 1984).

### 1.2.2 Phages of *Rhodococcus*

Two nocardiphages,  $\phi$ C and  $\phi$ EC, were isolated from soil and their integration sites mapped in the *Rhodococcus* chromosome (Brownell, 1976; Brownell and Adams, 1976; Brownell *et al.*, 1967). Host range studies showed that phage  $\phi$ C was specific for *R. canicruria* whereas phage  $\phi$ EC could be propagated on *R. canicruria*, *R. calcaria*, *R. globerula* and *R. rhodochrous* (Brownell *et al.*, 1967). These last four strains of *Rhodococcus* are all now regarded as strains of *R. erythropolis* (Goodfellow and Alderson, 1977) but which, nevertheless, show different genetic behaviour (Brownell, 1978).

The isolation of phages from rhodococci, as well as being useful in linkage studies, led to the characterisation of the phages themselves and investigations into their possible use as cloning vectors for *Rhodococcus* strains. The genome of phage  $\phi$ EC was estimated to be about 43.2 kbp by electron microscopy and a restriction endonuclease map suggested a size of about 46.0 kbp (Brownell and Denniston-Thompson, 1980). Phage  $\phi$ EC is a temperate phage which was modified to lysogenize the host *R. erythropolis* at a high frequency (Crockett and Brownell, 1972). The integration site of actinophage  $\phi$ EC was mapped in a lysogenic strain of *R. erythropolis* (Brownell and Adams, 1976). In addition, some genetic evidence was presented that the  $\phi$ EC prophage may exist as a plasmid as well as a chromosomally

integrated element (Brownell and Adams, 1976). These properties of  $\phi$ EC suggested that it might be a suitable candidate for development as a *Rhodococcus*-specific cloning vector. A deletion mutant of 37 kbp of  $\phi$ EC was made which contained potential cloning sites and which, it was hoped, would be able to accommodate foreign DNA (Brownell *et al.*, 1982). Techniques for transfecting DNA from phage  $\phi$ EC into *R. erythropolis* protoplasts exist and efficient protoplasting procedures required for the transfection and transformation of *Rhodococcus* cells have been described (Brownell, 1981; Brownell *et al.*, 1982) but transduction of *R. erythropolis* chromosomal markers has not been achieved with any of the derivatives of phage  $\phi$ EC.

A generalised transducing phage for *R. erythropolis*, Q4, was isolated from "soil around the roots of vegetation growing on sand dunes at the mouth of the Qolora River, Transkei" (Dabbs, 1987). Phage Q4 was found to be lysogenic and transduction to prototrophy of three unlinked auxotrophic loci as well as several antibiotic resistance loci was demonstrated with this phage. In addition, linkage was demonstrated between the loci for resistance to thiostrepton and rifampicin by these transduction studies. Segregation of these two markers was observed and phage Q4 was used to order the loci in a three factor cross, tetracycline resistance being the third factor (Dabbs, 1987). The generalised transducing bacteriophage Q4 may prove to be very useful in fine-structural genetic mapping of *Rhodococcus* species although transduction has only been demonstrated in two species, *R. erythropolis* and *Nocardia restricta*. In addition, the frequency of transduction in *R. erythropolis* was very low ( $10^{-8}$  pfu) although altered growth conditions or mutants may improve the host range and the transduction frequency (Dabbs, 1987).

### 1.2.3 Naturally occurring plasmids

A limited number of naturally occurring *Rhodococcus* plasmids have been described. As previously mentioned, the actinophage  $\phi$ EC can behave as a plasmid in *R. erythropolis* (Crockett and Brownell, 1972). In addition, native plasmids were found in the hydrogen-oxidizing autotrophic strain *Rhodococcus* sp. (*Nocardia opaca* 1b) (Reh and Schlegel, 1981; Sensfuss *et al.*, 1986). The phenotype for hydrogen autotrophy, known as the Aut character, enables bacteria to use gaseous hydrogen as the sole electron donor. It was shown that the Aut character, which consists of a set of at



least eight genes, was transmitted by conjugation during mating experiments with Aut<sup>-</sup> strains of *Rhodococcus*. Initial evidence pointed to the set of enzymes which determine the Aut character being plasmid encoded (Reh and Schlegel, 1981). However, further studies of transconjugants revealed that in many cases the Aut marker was transferred without plasmid transfer and that plasmid-free, Aut<sup>+</sup> transconjugants functioned as donors of the Aut marker. The Aut character was thus assumed to reside on the chromosome, functioning as an independent, self-transmissible, genetic element (possibly a transposon) and the plasmid was assumed to be cryptic (Sensfuss *et al.*, 1986). However, during the course of this same study, a marker for resistance to thallium was shown to be plasmid encoded and transferrable by conjugation.

A small (2.7 kbp) cryptic plasmid, pKU100, was isolated from *R. corallinus* (*Nocardia corallina*) and characterised (Kirby and Usdin, 1985). A number of unique restriction endonuclease sites suitable for cloning purposes was located and the plasmid was estimated to have a high copy number. The properties of this plasmid were considered to be suitable for a *Rhodococcus* cloning vector.

The lack of selectable resistance plasmids hampered the effort to develop *Rhodococcus* cloning vectors. However, a set of very large plasmids bearing resistance genes to the heavy metal compounds sodium arsenate, sodium arsenite and cadmium chloride and the antibiotic chloramphenicol were isolated from a strain of *R. erythropolis* which, it was hoped, might help to overcome this problem (Dabbs and Sole, 1988). The resistance genes appeared to be of chromosomal origin suggesting that plasmid integration into the genome had occurred. The plasmids, inexplicably, contained parts of the nocardiphage Q4 genome which had been used in the plasmid isolation procedure, and the plasmids were so large that they migrated more slowly than chromosomal DNA in agarose gels (Dabbs and Sole, 1988), adverse characteristics for potential cloning vectors although the marker genes appeared to be useful.

Another selectable marker was detected on a plasmid in *Rhodococcus fascians* (*Corynebacterium fascians*), a pathogen of dicotyledenous (Stapp, 1981) and monocotyledenous plants (Faivre-Amiot, 1967; Miller, 1980; Vantomme *et al.*, 1982). Previous reports had presented equivocal evidence for plasmid encoded virulence in different strains of *R. fascians* (Lawson *et al.*, 1982; Murai, 1981; Murai *et al.*, 1980).

A 138 kbp plasmid, pD188, was discovered in a strain of *R. fascians* encoding cadmium resistance which was transferrable between *R. fascians* strains by a "conjugation-like" mating system (Desomer *et al.*, 1988) which it was hoped would be useful for the genetic analysis of pathogenicity of *R. fascians*. However, neither this plasmid nor a range of other cadmium resistance and chloramphenicol resistance plasmids tested showed any correlation with virulence towards *Nicotiana tabacum* (Desomer *et al.*, 1987).

#### 1.2.4 New techniques and plasmids

High-voltage electroporation (Chassy *et al.*, 1988) using a range of engineered shuttle-cloning vectors that were capable of replicating in *R. fascians* and *E. coli* was applied successfully to the transformation of *R. fascians* with efficiencies ranging from  $10^5$  to  $10^7$  transformants per  $\mu\text{g}$  of DNA (Desomer *et al.*, 1990). These combined tools are being applied to the molecular analysis of interesting phenotypes of *R. fascians* and may find general application amongst other rhodococci.

An *E. coli*-*Rhodococcus* shuttle vector was developed which made stable transformants in a host range which included strains of *R. erythropolis*, *R. globerulus* and *R. equi* (Vogt Singer and Finnerty, 1988). Stable transformants were not obtained, however, with strains of *R. rhodochrous* or several strains of coryneform bacteria. A polyethylene glycol-assisted transformation method developed for this shuttle vector yielded more than  $10^5$  transformants per  $\mu\text{g}$  of DNA. This transformation system was used to express the thiostrepton resistance gene, *tsr*, from *Streptomyces* and the *E. coli* ampicillin resistance gene, *bla*, in *Rhodococcus*. This was the first report of a *Rhodococcus* plasmid transformation system and of heterologous gene expression in *Rhodococcus* spp.

*Rhodococcus* spp. are generally easier to grow under laboratory conditions than *Streptomyces* and the development of efficient systems for genetic manipulations for *Rhodococcus* may make it a useful organism for the cloning and study of other actinomycete genes such as those from pathogenic strains of *Mycobacteria* and *Nocardia*. Actinomycete genes often are poorly expressed in *E. coli* due to differences in promoter structure and promoter recognition factors (Bibb *et al.*, 1985; Bibb and Cohen, 1982). Their genetic regulatory systems may function better in *Rhodococcus*, another actinomycete. It is already established that mycobacterial genes are often

better expressed from their own promoters in *Streptomyces* strains than in *E. coli* (Clark-Curtiss *et al.*, 1985; Kieser *et al.*, 1986; Thole *et al.*, 1985).

### 1.2.5 Gene cloning studies

A pigment gene was cloned from a buff coloured *Rhodococcus* sp. The gene was expressed in *E. coli* and blue colonies were formed (Hill *et al.*, 1989). The nucleotide sequence of this gene has been published (Hart *et al.*, 1990).

The only other *Rhodococcus* gene cloned and sequenced to date was the nitrile hydratase gene of *Rhodococcus* species N-774 (Ikehata *et al.*, 1989). Nitrile hydratase transforms acrylonitrile into acrylamide and in addition possesses some unique enzymatic properties. This gene was isolated from a *Rhodococcus* gene library by hybridization to two synthetic oligonucleotide probes which were complementary to the amino-terminal amino acid sequences of the  $\alpha$  and  $\beta$  subunits of *Rhodococcus* N-774 nitrile hydratase. This enzyme was expressed in *E. coli* under control of the *lac* promoter of pUC19. The cloning and expression of *Rhodococcus* nitrile hydratase will facilitate the biochemical analysis of the enzyme and may be important in the industrial production of acrylamide.

It is clear that the genetics of *Rhodococcus* is still very much in its infancy although the development of the new transformation techniques and cloning vectors described above should be beneficial to our understanding and exploitation of this potentially, uniquely useful organism.

## 1.3 BACTERIAL PIGMENTS

Some of the earliest bacterial genetic studies were performed on pigmented bacteria as they provided a convenient, visible phenotype in an organism which had few others. The biochemical and genetic analysis of pigment production in many species of bacteria has elicited information not only on the biosynthesis of the pigments themselves but has also provided useful model systems for fundamental genetic studies. For example, studies on the molecular genetics of actinorhodin biosynthesis have helped in the elucidation of the structure and organization of antibiotic biosynthesis genes in streptomycetes (Malpartida and Hopwood, 1984;

1986) and production of the pigment prodigiosin was shown to be related to acriflavine resistance in the opportunistic pathogen *Serratia marcescens* (Feng *et al.*, 1982; Woods *et al.*, 1971; 1973).

### 1.3.1 Pigments of *Rhodococcus*

Many strains of *Rhodococcus* produce pigmented colonies of buff, pink, orange or red. The pigment of *Rhodococcus corallinus* was purified and separated into a major crystalline red compound which did not appear to be a carotenoid, and a yellow oil (Brown and Clark, 1966). NMR analysis indicated the presence of an aromatic ring, a long aliphatic hydrocarbon chain and carbonyl groups suggesting that the pigment belonged to the class of actinomycins (Stees *et al.*, 1969). More recently, however, four kinds of carotenoid pigments were isolated and identified from a strain of *R. rhodochrous*, which forms light orange coloured colonies (Takaichi *et al.*, 1990).

### 1.3.2 Pigments of *Streptomyces*

*Streptomyces coelicolor* produces the pH-sensitive, red-blue pigment actinorhodin which has antibiotic activity, and a red cell wall associated compound, undecylprodigiosin (Rudd and Hopwood, 1980). These two pigments are chemically distinct and are generally regarded as secondary metabolites. The genes determining the synthesis of these two pigments were identified (Feitelson *et al.*, 1985; Malpartida and Hopwood, 1986; Rudd and Hopwood, 1979; 1980) and cloned (Feitelson and Hopwood, 1983; Feitelson *et al.*, 1986; Malpartida and Hopwood, 1984; 1986). However, the physiological controls which regulate pigment production in *S. coelicolor* are unknown. Hobbs *et al.* (1990) have demonstrated that the kinetics of accumulation of undecylprodigiosin and actinorhodin are markedly disparate and that the composition of the growth medium affects the synthesis of both products. The genetic pathway of actinorhodin biosynthesis was genetically analysed and sequentially ordered by the use of blocked mutants (Cole *et al.*, 1987; Floss *et al.*, 1985; Rudd and Hopwood, 1979) and a biochemical pathway recently was proposed for the biosynthesis of actinorhodin (Bartel *et al.*, 1990). A part of the actinorhodin biosynthetic pathway on a plasmid was used to develop a chromogenic promoter-probe vector for the identification of *Streptomyces* promoters (Horinouchi and Beppu, 1985). This plasmid encoded production of a brown

pigment in *Streptomyces lividans*, which was assumed to be a shunt product of the actinorhodin biosynthetic pathway, when a promoter was inserted upstream of the pigment producing genes.

The production of melanin is widespread amongst the streptomycetes. Tyrosinase, a copper-containing mono-oxygenase, converts the amino acid tyrosine to melanin in a single step. The structural gene for tyrosinase, *melC2*, was cloned (Katz *et al.*, 1983) and shown to be cotranscribed with an upstream open reading frame, *melC1* (Bernan *et al.*, 1985). Tyrosinase has a requirement for copper and *melC1* probably encodes a copper-transferring protein (Lee *et al.*, 1988).

A novel tyrosine-derived pigment with anti-dermatophyte activity, designated SL-1 was isolated from a mutant of *S. lavendulae* (Mikani *et al.*, 1987). This pigment was reddish brown in colonies on solid media and wine red in liquid media. It was yellow or orange in acid to neutral solutions and wine red to redish brown in alkaline solutions. A *Streptomyces* species was observed to produce a blue pigment, virginiamycin, under the control of an autoregulator (Yanagimoto *et al.*, 1988).

### 1.3.3 Pigment production by other bacteria

Bacterial pigments occur in various genera of both Gram-negative and Gram-positive bacteria. The pigments of Gram-positive cocci are primarily carotenoids; for example *Staphylococcus aureus* was shown to contain numerous carotenoid pigments (Sobin and Stahly, 1942) and the carotenoid pigment canthaxanthin was isolated from *Micrococcus roseus* (Cooney *et al.*, 1966).

Among the Gram-negative bacteria, prodigiosin is produced by *Vibrio psychroerythreus* (D'Aoust and Gerber, 1974) and *Pseudomonas mangesiorubra* (Gandhi *et al.*, 1973). Many *Pseudomonas* spp. are prolific producers of pigments such as the extracellular, water-soluble, yellow-green pigments produced by a range of fluorescent pseudomonads (Loper *et al.*, 1984). At least four genes or gene clusters were shown to be involved in the production of the fluorescent pigment of *P. syringae* pv. *syringae* (Loper *et al.*, 1984). The bright red-violet pigment methylosin is produced by the methanotropic bacterium *Methylosinus trichosporium* (Strauss *et al.*, 1983). Flexirubins are produced by members of the genus *Flavobacterium* (Shewan and McMeekin, 1983). Prodigiosin is produced by *S. marcescens* (Williams *et al.*,

1956) and was shown to consist of a blue fraction and three red fractions which had tripyrrole structures characteristic of prodigiosins (Green *et al.*, 1956).

*Erwinia herbicola* is a yellow pigmented member of the enterobacteriaceae that occurs widely as an epiphyte on plant surfaces (Billing and Baker, 1963; Goodman, 1965). The concomitant loss of pigment producing ability and thiamine prototrophy was observed after incubation at elevated temperatures or in the presence of nalidixic acid or SDS (Chatterjee and Gibbins, 1971; Thiry, 1984) and evidence was presented that the genes for pigmentation and thiamine prototrophy were encoded on a large plasmid (Gantotti and Beer, 1982). The genes involved in yellow pigmentation were cloned and expressed in *E. coli* (Perry *et al.*, 1986) and evidence indicated that the yellow pigments were carotenoids (Sandmann *et al.*, 1990; Tuveson *et al.*, 1988).

#### 1.3.4 Pigment production by *E. coli*

*E. coli* does not normally produce pigments but a number of pigment producing systems have been cloned into *E. coli* and expressed, enabling further investigation of the biochemistry and genetics of these systems and the development of plasmids containing the pigments as markers.

DNA sequences from *S. marcescens* encoding parts of the prodigiosin biosynthetic pathway were cloned into *E. coli* which was induced to produce the bright red pigment prodigiosin when the appropriate substrates were present in the growth media (Dauenhauer *et al.*, 1984).

The genes encoding yellow carotenoid pigment production in *E. herbicola* were cloned in *E. coli* and localised to a 12.4 kbp chromosomal fragment (Perry *et al.*, 1986). At least seven poly-peptides were produced by this 12.4 kbp fragment and the expression of the yellow pigment phenotype in both *E. herbicola* and *E. coli* was subject to catabolite repression. DNA hybridization studies indicated that different yellow pigment genes exist among different *E. herbicola* strains. Carotenoid pigments are known to protect against damage by near UV in photosynthetic bacteria (Fujimori and Livingston, 1957; Griffiths *et al.*, 1955), *Myxococcus xanthus* (Burchard and Dworkin, 1966; Burchard *et al.*, 1966) and possibly nonphotosynthetic bacteria (Mathews and Sistrom, 1963; 1959). The genes controlling carotenoid biosynthesis from *E. herbicola* were cloned into *E. coli* to investigate whether there were any protective effects of these yellow pigments of *E. herbicola* against reactive

oxygen species, which are produced after absorption of visible light by chlorophyll present in the host plant (Tuveson *et al.*, 1988). Pigment producing *E. coli* cells were significantly protected against three photosensitizing molecules activated by near UV ( $\alpha$ -terthienyl, harmine and phenylheptatriyne) which were thought to have a lethal affect on the bacterial membrane. The yellow pigments of *E. herbicola* expressed in *E. coli* were identified as carotenoids and as the protecting factor against near UV in the presence of  $\alpha$ -terthienyl (Sandmann *et al.*, 1990).

The *Streptomyces antibioticus* melanin-producing *melC* operon (1.3 kbp) was transferred to an *E. coli* plasmid in such a way that its transcription was under the control of the *lac* promoter and was translationally fused to the *lacZ*  $\alpha$ -fragment which contained a cloning cassette. *E. coli* cells containing this plasmid produced melanin after overnight incubation on agar supplemented with 0.1 mM  $\text{CuCl}_2$ , 0.36 mM IPTG and 2 mM tyrosine (Altenbuchner, 1988; Tseng *et al.*, 1990) and provided a convenient cloning marker for insertional or replacement inactivation.

Another melanin producing vector was made which could use various tyrosine analogues as substrates producing a range of pigments from yellow to various shades of brown (della-Cioppa *et al.*, 1990). The melanin operon of *S. antibioticus* which was cloned and expressed in *E. coli* had the potential for the development of new melanin based pigments which have the ability to act as UV absorbers, cation exchangers, amorphous semiconductors and novel biopolymers with drug binding and possibly other properties (Bell and Wheeler, 1986).

## 1.4 THE PRODUCTION OF INDIGO

### 1.4.1 Historical

Indigo is an important and valuable dyestuff with a long history of use by man. It was used widely by the ancient civilisations of Egypt, Greece and Rome, being easily obtained from plants of the genus *Indigofera*. Julius Caesar observed its use by the Britons who obtained it from the woad plant, *Isatis tinctoria*. Indigo continues to be used widely in the textile industry for the dyeing of cotton and wool.

Indigo occurs as a colourless glucoside called indican,  $\text{C}_{14}\text{H}_{17}\text{O}_6\text{N}$ , in various species of *Indigofera* including *sumatrana*, *arrecta*, *tinctoria*, *pseudotinctoria*, *anil*,

*disperma* and *argentea*. These plant species were originally indigenous to the region between latitudes 20° and 30° of Eastern Asia but they also thrive in China, Japan, the Philippines, Central America, Brazil and Java.

Common woad, *I. tinctoria*, was formerly cultivated in Europe but yields a smaller amount of indican than *Indigofera*. The cultivation of *Indigofera* and the extraction of the dyestuff was an important industry in India up to the beginning of the 20<sup>th</sup> century. In 1897 1,750,000 acres of Indian land were under *Indigofera* cultivation and approximately 17,000,000 lb of indigo were exported. The cultivation of *Indigofera* was introduced successfully to the Carolinas of North America in 1744 and was economically important to the South in the American civil war.

#### 1.4.2 Preparation and synthesis

The method of preparing indigo from *Indigofera* has remained essentially unchanged for centuries. The dye is prepared from freshly cut plants which are macerated, packed into large vats and allowed to ferment. In this process the water-soluble indican is hydrolyzed to glucose and indoxyl. By mild oxidation, even by air, two molecules of indoxyl are converted by the loss of four hydrogen atoms into indigo,  $C_{16}H_{10}O_2N_2$ , a dark blue powder. The indigo precipitate is strained, pressed and dried into cakes. For use as a paint pigment it is not precipitated with a mordant but ground directly to a fine powder. In the textile industry, indigo is used as a vat dye, i.e. it is insoluble in water and must be converted to a soluble product (indigo white) by alkaline reduction in order to prepare the dye-vat. After steeping the fabric in the dye-vat, insoluble indigo is deposited on the fibre by aerial oxidation.

The chemical structure of indigo was announced by the German chemist A. von Bayer in 1883 although it was not until 1898 that an economically sound process was developed for the commercial production of synthetic indigo. The original commercial synthesis of indoxyl was initiated by Badische Anilin und Soda Fabrik, Germany, but it was developed in the plant of Meister, Lucius and Breuning at Hoechst. The reaction essentially consisted of the fusion of phenylglycine sodium salt,  $C_6H_5.NH.CH_2COONa$ , in a mixture of caustic soda and sodium amide to produce indoxyl. The use of sodium amide was the key to the success of the development because it allowed the fusion to take place at a lower temperature, thereby improving the yield. The introduction of relatively cheap synthetic indigo



caused an economic upheaval in India at the beginning of the 20<sup>th</sup> century and the cultivation of indigo-bearing plants in India has now practically ceased. Since World War II the production of synthetic indigo has spread to every country which possesses a dyestuff industry.

Indigo is a blue pigment which develops a coppery tint when rubbed and has a fair tinting strength although it may fade rapidly when thin and exposed to strong sunlight; yet specimens of it many centuries old have been discovered without apparent change. Many substituted derivatives of indigo are known which have improved properties as dyestuffs including 6,6-dibromoindigo which appears to be identical with the "Purple of the Ancients".

#### 1.4.3 Bacteria and indigo

The bacterial production of a blue pigment by three species of bacteria grown on media containing indole was described in 1927 (Gray, 1927). This blue pigment was identified as indigo by chemical tests and absorption spectrum analysis. One of the bacterial species was identified as *Pseudomonas indoloxidans* and the others as *Micrococcus piltonensis* and *Mycobacterium globerulum*. The description of *M. globerulum* (Gray, 1927) applied to more recent taxonomic systems suggests that it may belong to the genus *Rhodococcus* (Goodfellow and Alderson, 1977). The production of indigo by *P. indoloxidans* was studied as this organism produced far more pigment than either of the other two species. It was shown that *P. indoloxidans* oxidized indole to indigo in liquid and solid media but that tryptophan, which is an amino acid derivative of indole, did not act as a substrate for indigo production. The indigo was formed as crystals outside the organism and occurred only in the growth phase of *P. indoloxidans* cultures. There was very little indigo production when indole was added to the culture in stationary phase. The oxidation of indole to indigo by *P. indoloxidans* was further investigated biochemically (Oshima *et al.*, 1965). They found that two molecules of indole were oxidized to one molecule of indigo by the organism; that indoxyl was an intermediate in the reaction and that the incorporation of oxygen atoms from molecular oxygen into indigo was observed in the course of the oxidation of indole by the organism. These authors suggested that an enzyme, indole-3-hydroxylase, was responsible for the transformation of indole to indigo via indoxyl which was oxidized in air to indigo. Some indole

derivatives (indole 6-carboxylic acid, 6-hydroxyindole, indole-3-carboxylic acid, 3-methylindole and 2,3-dimethylindole) were tested as substrates but none were capable of indigo formation suggesting that the enzyme was very specific for indole. A later study on bacterial oxidation of indole (Fujioka and Wada, 1968) reported the isolation of an unidentified Gram-positive coccus from the soil which rapidly decomposed indole and produced copious amounts of a blue pigment. These authors presented evidence that indole was metabolized to anthranilic acid via dihydroxyindole. Although they were unable to isolate an enzyme which metabolized indole itself, they were able to isolate an enzyme which oxidized dihydroxyindole to anthranilic acid and was induced by indole. They did not analyse the blue pigment although its formation in association with indole oxidation suggests it may have been indigo. The production of two pigments, indigo and violacein, in divergent pathways of indole metabolism was observed in the bacterium *Chromobacterium violaceum* (Sebek and Jager, 1967).

#### 1.4.4 Aromatic dioxygenases and indigo production from recombinant *E. coli* clones

The cloning and expression of genes from *Pseudomonas putida* led to the serendipitous discovery of indigo production by recombinant *E. coli* clones and a clearer understanding of the pathway of indigo formation from indole (Ensley *et al.*, 1983). *P. putida* contains a plasmid pNAH7 which encodes genes for the conversion of naphthalene to salicylic acid, a process of potential pharmaceutical and commercial interest. The first step in experiments to analyse the pathway involved cloning fragments of pNAH7 in pBR322 into *E. coli*. It was shown that the entire pathway for the formation of salicylic acid from naphthalene had been cloned and, quite unexpectedly, that one of the recombinants was producing a dark blue pigment that on further analysis was identified as indigo. The criteria for identifying the pigment as indigo were its solubility characteristics, mobility on thin-layer chromatography and visible, ultra-violet and mass spectra. Analysis of the pigment-producing recombinant *E. coli* strain suggested that the metabolic interactions that produced indigo involved the pNAH7 encoded naphthalene dioxygenase and indole. Naphthalene dioxygenase is a three-component enzyme system which converts naphthalene into *cis*-naphthalene dihydrodiol (Ensley *et al.*,

1982). Indole is constitutively produced by *E. coli* as part of the normal metabolism of the cell in the degradation of tryptophan by tryptophanase. Evidence supporting the production of indigo from indole included the observation that the supplementation of the growth media with 10 mM tryptophan or 1 mM indole enhanced pigment production in the recombinant, and that indigo formation was not observed in cultures containing Luria broth and one per cent glucose. High glucose concentrations can cause catabolite repression of tryptophanase synthesis in *E. coli* (Bilezikian *et al.*, 1967; Freundlich and Lichstein, 1960), thus limiting the production of indole. *P. putida* appeared to lack tryptophanase activity as it produced indigo only when indole was supplied in the media. Analysis of *P. putida* mutant strains provided further evidence that naphthalene dioxygenase catalyzed the formation of indigo from indole. Evidence that indole oxidation was a general property of bacterial dioxygenases was also presented (Ensley *et al.*, 1983) and a possible pathway for indigo production was described which also explained previous accounts of indigo production. In this pathway indole was oxidized by naphthalene dioxygenase to *cis*-indole-2,3-dihydrodiol from which one molecule of water was spontaneously eliminated to form indoxyl. Indigo was formed by the oxidation in air of indoxyl.

The formation of indigo by a bacterial aromatic dioxygenase from indole suggested that this could be a general property of this class of enzymes and a useful tool for the elucidation of their mechanism of action (Ensley *et al.*, 1983). The observation of indigo production by recombinant *E. coli* clones was used to isolate a number of other aromatic dioxygenase genes including those for benzene, isopropylbenzene and toluene and to examine their roles in their particular pathways. In each case the production of indigo could be explained in terms of a dihydrodiol forming ability. The observation of indigo-forming *E. coli* clones was used to isolate the gene for isopropylbenzene dioxygenase from *P. putida* RE204 (Eaton and Timmis, 1986). *P. putida* RE204 is capable of utilizing iso-propylbenzene as its sole carbon and energy source. By transposon Tn5 mutagenesis the pathway was localised to a plasmid pRE4 which was then cloned in fragments in *E. coli*. It was shown that the entire pathway had been cloned and the order of the pathway was established. *E. coli* clones containing the gene for isopropylbenzene dioxygenase which catalyses the first step of the pathway, (isopropylbenzene to 2,3-dihydro-2,3-

dihydroxyisopropylbenzene), were easily identified by their ability to produce indigo on indole-containing plates.

*P. putida* BE-81 is capable of assimilating benzene in a pathway which converts benzene to *cis*-benzene glycol and subsequently to catechol (Gibson *et al.*, 1968). The first step of this pathway is catalysed by benzene dioxygenase. The pathway was cloned in a single fragment and was shown to produce a green pigment which was enhanced by addition of indole or tryptophan to the growth media and was inhibited by the addition of glucose to the growth media, leading to the assumption that the pigment was indigo (Irie *et al.*, 1987a). The cloned benzene assimilating genes were subsequently analysed by nucleotide sequencing and the benzene dioxygenase enzyme more fully characterised (Irie *et al.*, 1987b). The genes encoding this pathway were thought to be chromosomally encoded. In contrast to this finding, the gene for benzene dioxygenase in *P. putida* ML2 was shown to be encoded on a large (112 kbp) plasmid (Tan and Mason, 1990), fragments of which were cloned in *E. coli* and the expressed benzene dioxygenase gene detected by the production of indigo. This was confirmed by Western blotting analysis using an antibody raised to benzene dioxygenase.

*Pseudomonas putida* F1 can utilise toluene as the sole carbon and energy source by the toluene dioxygenase enzyme system (Yeh *et al.*, 1977). The initial step of this pathway is catalysed by toluene dioxygenase which converts toluene into *cis*-toluene dihydrodiol through the addition of two atoms of molecular oxygen to the aromatic nucleus (Gibson *et al.*, 1970; Kobal *et al.*, 1973; Ziffer *et al.*, 1973). The production of indigo was used as a marker to indicate the presence of an intact toluene dioxygenase system (Zylstra *et al.*, 1988). The genes encoding this pathway were cloned from chromosomal DNA and sequenced (Zylstra and Gibson, 1989). The nucleotide sequence of toluene dioxygenase was compared with that of benzene dioxygenase and they were shown to be almost identical over a length of 5,292 bp. In addition, benzene dioxygenase was shown to be immunologically homologous with toluene dioxygenase but not naphthalene dioxygenase (Zamanian and Mason, 1987).

The use of bacterial fermentation systems to produce indigo on a commercial basis using indole as the substrate was investigated (Murdock, Sendar and Ensley, 1990). The limiting factors in the process were the short half-life of the naphthalene

dioxygenase enzyme system and the high cost of indole. The unstable component of the naphthalene dioxygenase enzyme system was shown to be ferredoxin<sub>NAP</sub>. Enzyme activity was increased by the addition of iron to the fermentation medium and increasing the concentration of ferredoxin<sub>NAP</sub> by gene amplification and by altering the ferredoxin<sub>NAP</sub> gene by site-directed mutagenesis. These measures resulted in the enzyme half-life being increased from 1.5 h to 15 h. Site-directed mutagenesis was also used to increase the bacterial synthesis of tryptophan. These mutants were able to produce indole instead of tryptophan from glucose. The combined effect of altering the tryptophan metabolism and the naphthalene dioxygenase enzyme system was the rapid biosynthesis of indigo from glucose. The biosynthesis of indigo from genetically engineered microorganisms may lead to the cheaper production of this pigment.

## CHAPTER TWO

### ANALYSIS OF THE PIGMENTS AND THE PATHWAY OF PIGMENT PRODUCTION

|  |    |
|--|----|
| 2.1 INTRODUCTION   | 23 |
| 2.2 METHODS AND MATERIALS  | 24 |
| 2.2.1 Bacterial strains, plasmids and media                                | 24 |
| 2.2.2 Identification of pigment producing clones                           | 24 |
| 2.2.3 Extraction of pigments from liquid cultures                          | 24 |
| 2.2.4 Thin layer chromatography  | 25 |
| 2.2.5 Absorption spectra of the pigments                                   | 26 |
| 2.2.6 Solubility of the pigments   | 26 |
| 2.2.7 Transformation of mutant strains                                     | 27 |
| 2.2.8 <i>Rhodococcus</i> growth experiments                                | 27 |
| 2.2.9 <i>In vitro</i> assay for pigment formation                          | 27 |
| 2.2.10 Minimal media supplementation                                       | 27 |
| 2.2.11 RNA extraction  | 28 |
| 2.2.12 RNA dot blot analysis   | 28 |
| 2.3 RESULTS  | 29 |
| 2.3.1 Minimal media supplementation  | 29 |
| 2.3.2 Transformation of a tryptophanase mutant strain<br>of <i>E. coli</i> | 31 |
| 2.3.3 <i>Rhodococcus</i> growth experiments                                | 33 |
| 2.3.4 <i>In vitro</i> assay for pigment production                         | 33 |
| 2.3.5 Thin layer chromatography  | 33 |
| 2.3.6 Pigment absorption spectra   | 34 |
| 2.3.7 Chemical properties of the pigments                                  | 34 |
| 2.3.8 NMR spectroscopy   | 38 |
| 2.3.9 RNA dot blot analysis  | 38 |
| 2.4 DISCUSSION   | 40 |

## CHAPTER TWO

### ANALYSIS OF THE PIGMENTS AND THE PATHWAY OF PIGMENT PRODUCTION

**SUMMARY:** Pigment was produced by *E. coli* containing a cloned fragment of DNA from *Rhodococcus* sp. ATCC 21145. Crude chloroform extracts of the pigment were separated into blue and pink components and it was shown that the pink component was derived from the blue component. Evidence from thin layer chromatography, NMR spectroscopy, absorption spectrum analysis and solvent solubility behaviour suggested that the blue pigment was indigo and the pink pigment was indirubin, a structural isomer of indigo. Tryptophan and indole were precursor substrates for pigment production in *E. coli* containing plasmids with the *Rhodococcus* pigment-producing DNA insert. The proposed pathway for pigment production on LB agar involves the conversion of tryptophan to indole by tryptophanase of *E. coli* and the oxidation of indole to indigo by the product of the cloned *Rhodococcus* DNA insert. This was supported by the observation that pigment production was greatly reduced in a tryptophanase mutant strain of *E. coli*. Indigo production by *E. coli* containing the cloned genes for several pseudomonad aromatic dioxygenases has been described previously and this phenotype has been used to detect aromatic dioxygenase expression, suggesting that the enzyme encoded by the *Rhodococcus* DNA insert also belonged to this family of enzymes.

## 2.1 INTRODUCTION

The aims of the studies presented in this chapter were to identify the pigments and the pathway by which they were produced in *E. coli* containing the pigment-producing DNA insert of *Rhodococcus* sp. ATCC 21145. The elucidation of the nature of these pigments would facilitate the optimization of bacterial pigment biosynthesis which was required for the development of an efficient reporter gene in an insertional-inactivation cloning vector, and for investigations into the possible uses and applications of the pigment itself.

Initial pigment studies showed that the pink and blue components of chloroform extracts of the pigmented *E. coli* strains were soluble at low concentrations in chloroform and insoluble in water, ethanol and methanol (Hill *et al.*, 1989). The compounds flocculated at concentrations above 20 mg ml<sup>-1</sup> in chloroform and did not change colour on the addition of acids or alkalis. These pigments appeared to be different to the prodigiosins and actinorhodins commonly produced by actinomycetes.



## 2.2 METHODS AND MATERIALS

### 2.2.1 Bacterial strains, plasmids and media

The bacterial strains used in these studies, listed in Table 2.1, were maintained routinely on LB agar (Appendix B). Plasmids in *E. coli* transformants were selected by supplementing media with the antibiotic Ap (100 µg ml<sup>-1</sup>). Lactose (lac) was added to MM at a final concentration of 0.2 mg ml<sup>-1</sup>. Tryptophan and indole were added to media at a final molarity of 1 mM. Tryptophan and indole were alternatively added to the surfaces of plates by spreading 100 µl of 20 % and 10 % solutions of each respectively and allowing them to dry for several hours before streaking bacterial cells onto the plates.

### 2.2.2 Identification of pigment producing clones

Pigment producing *E. coli* clones were readily apparent on agar media as blue to blue-grey colonies after overnight incubation at 37 °C followed by 8-24 h incubation at room temperature.

### 2.2.3 Extraction of pigments from liquid cultures

The following pigment extraction procedure was applied to liquid cultures of *E. coli* LK111 (pNC185), *E. coli* LK111 (pUC18), *Rhodococcus* sp. ATCC 21145 and *Rhodococcus* sp. JL10. Bacteria were inoculated into 400 ml LB with Ap if necessary and incubated overnight at 37 °C with good aeration. Cells were harvested by centrifugation at 12,000 × *g* for 10 min and washed once with distilled water before resuspending the cells in a small volume (less than 10 ml) of water. The suspension of cells was then disrupted by sonication and 100 ml of chloroform (analytical grade) was added and shaken vigorously at 37 °C for 12-16 h. The efficiency of pigment extraction was improved by changing the chloroform twice which involved pelleting the cell debris by centrifugation at 12,000 × *g* for 10 min and resuspending in fresh chloroform before continuing the vigorous shaking. After 16 h the blue chloroform extracts were combined and evaporated overnight in a fume hood and the pigment redissolved in a minimal volume of fresh chloroform.

Table 2.1 Strains and plasmids.

| <u>STRAIN</u>                        | <u>GENOTYPE</u>   | <u>SOURCE</u>                 |
|--------------------------------------|---|-------------------------------|
| <i>E.coli</i> 3.300                  | <i>lacI</i> 22, $\lambda^-$ ,<br><i>relA1</i> , <i>spoT1</i> , <i>thi-1</i> | Pardee <i>et al.</i> (1959)   |
| <i>E. coli</i> LK111                 | <i>lacI</i> , <i>lacZ</i> , $\Delta$ M15<br><i>lacY</i> <sup>+</sup>        | Zabeau and Stanley,<br>(1982) |
| <i>E. coli</i> MY252                 | <i>tna</i> <sup>+</sup> , $\Delta$ <i>trpAC9</i>                            | M. Yudkin, Oxford, UK         |
| <i>E. coli</i> MY1393                | <i>tna</i> 103, $\Delta$ <i>trpAC9</i>                                      | M. Yudkin, Oxford, UK         |
| <i>Rhodococcus</i> sp.<br>ATCC 21145 | wild type   | ATCC 21145                    |
| <i>Rhodococcus</i> sp.<br>JL10       | wild type   | Hill <i>et al.</i> (1989)     |
| <u>PLASMIDS</u>                      | <u>PHENOTYPE</u>  |                               |
| pKS1                                 | Ap <sup>r</sup> LacZ <sup>-</sup> Pig <sup>+</sup>                          | This study                    |
| pKS3                                 | Ap <sup>r</sup> LacZ <sup>-</sup> Pig <sup>+</sup>                          | This study                    |
| pNC185                               | Ap <sup>r</sup> LacZ <sup>-</sup> Pig <sup>+</sup>                          | This study                    |
| pNC181                               | Ap <sup>r</sup> LacZ <sup>-</sup> Pig <sup>+</sup>                          | This study                    |
| pUC18                                | Ap <sup>r</sup> LacZ <sup>+</sup>   | Vieira & Messing (1982)       |

#### 2.2.4 Thin layer chromatography

Preparative thin layer chromatography (TLC) was performed to separate the pink and blue components of the crude, pigmented chloroform extracts from each other as well as other bacterial, chloroform-soluble contaminants. Chloroform was used as the chromatographic solvent. Samples of the pigmented chloroform extract were applied repeatedly across the width of the plate (Merck cat. no. 5717, PLC silica gel 60 F<sub>254</sub>, 20 cm x 20 cm, 2 mm thick). Each application was allowed to dry completely before the next one was applied. The chromatographs were resolved by standing the plates in a glass chromatographic tank containing a saturated atmosphere of chloroform and a small volume of chloroform so that the solvent was in contact with

the silica stationary phase. The crude pigment extract separated into pink and blue pigmented bands within 40 min. After drying the plates, the silica containing the pigment bands was cut and scraped from the TLC plate, crushed to a fine powder and poured into a glass bottle containing chloroform. The silica slurry containing the pigments was shaken vigorously for 1 h to allow the pigments to elute into the chloroform. The silica was removed by filtration in a Buchner funnel, washed with fresh chloroform and the purified pigment concentrated by evaporation of the combined chloroform washings. Purified pigment fractions were dissolved in DMSO for analysis by NMR (nuclear magnetic resonance) spectroscopy.

Analytical TLC was performed on chromatographic plates that had a much thinner silica coating (Merck cat. no. 16485, silica gel 60 F<sub>254</sub>, 20 cm x 20 cm, 0.25 mm thick). Samples were applied as small spots along a reference line and commercial indigo (Sigma) in chloroform (0.5 mg ml<sup>-1</sup>) was used as a control. *R<sub>F</sub>* values were determined by dividing the distance moved by a particular spot to the distance moved by the solvent front.

#### 2.2.5 Absorption spectra of the pigments

The visible absorption spectra (400-800 nm) of TLC-purified and unpurified bacterial pigment samples and an indigo control sample were determined by scanning in a Beckman DU-40 spectrophotometer attached to an Epson FX-80 printer.

#### 2.2.6 Solubility of the pigments

Freshly prepared bacterial pigment extracts in chloroform were evaporated to dryness, resuspended in a range of organic solvents (acetone, chloroform, 1,2-dichloroethane, dimethyl sulfoxide, ethanol, ethyl acetate, ethylmethylketone, glacial acetic acid, methanol, octane, toluene and xylene) and their solubility and appearance compared to that of commercial indigo.

### 2.2.7 Transformation of mutant strains

The *E. coli* tryptophanase mutant strain MY1393 and its wild type parent strain MY252 (Table 2.1) were both transformed with the plasmids pNC185 and pUC18 by the standard procedures (Appendix A).

### 2.2.8 *Rhodococcus* growth experiments

*Rhodococcus* sp. ATCC 21145 was grown on LB agar (Appendix A) and LB agar supplemented with either 1 mM indole (Sigma) or 1 mM tryptophan. Plates were incubated at 30 °C for up to five days and observed for pigment production.

### 2.2.9 *In vitro* assay for pigment formation

*E. coli* LK111 (pNC185) was inoculated into 20 ml LB Ap and incubated at 37 °C until the culture was in log phase. Cells were harvested by centrifugation at 3,000 x g for 5 min at 4 °C and washed in ice cold saline before resuspending in 10 ml of 50 mM KH<sub>2</sub>PO<sub>4</sub>/NaOH buffer, pH 7.5. The cells were disrupted with three 30 s bursts from a sonicator and the debris was pelleted by centrifugation of 1 ml samples of the suspension in a microfuge at 16,000 x g for 5 min at 4 °C. The clear cell-free supernatant was stored at -70 °C. *In vitro* pigment formation was assayed by adding 1 mM indole to the cell-free extract and incubating the reaction mixture at 37 °C for up to 24 h with occasional mixing and gentle shaking. The reaction mixture was extracted twice with equal volumes of ethyl acetate which were combined and the absorption spectrum determined in the range 400-800 nm.

### 2.2.10 Minimal media supplementation

*E. coli* 3.300 containing pNC185 was grown on MM lac Ap agar plates (Appendix A) supplemented by spreading 100 µl of a 20 % solution of one of the following amino acids on the surface of the plates: L-alanine, L-arginine, cysteine-HCl, L-glutamic acid, L-glycine, L-histidine, L-leucine, L-methionine, D,L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine or D,L-valine. The plates were allowed to dry for a few hours before inoculation and incubation at 37 °C for 48 h.

### 2.2.11 RNA extraction

The RNA extraction method was based on that of Aiba *et al.* (1981) and was used to extract high molecular weight RNA from both *E. coli* and *Rhodococcus* strains. Precautions were taken to prevent RNase contamination at all steps of the RNA isolation procedure such as treatment of glassware and solutions with 0.1 % DEPC (diethylpyrocarbonate) and the wearing of gloves.

Cells from which RNA was required were inoculated into an overnight 5 ml LB or LB Ap culture which was used to inoculate 100 ml LB or LB Ap. Cells were harvested at mid-log phase ( $OD_{600} = 0.5-0.6$ ) and resuspended in 3 ml of lysis buffer (0.1 mM Tris-HCl, pH 8.0, 1.0 mM EDTA, 1.0 % SDS) previously warmed to 60 °C. Three ml of hot phenol (60 °C) was added to the cell suspension, mixed well and held at 60 °C for 5 min with gentle shaking. The suspension was pelleted by centrifugation at 16,000 x g in a microfuge and the aqueous phase extracted with an equal volume of room temperature phenol then extracted twice with an equal volume of water saturated ether. RNA was precipitated from the aqueous phase at -20 °C after adding one tenth volume of 3 M sodium acetate and three volumes of absolute ethanol. Pellets were dissolved in lysis buffer, reprecipitated before resuspension in DNase buffer (100 mM sodium acetate, pH 5.0, 5.0 mM  $MgSO_4$ ) and pooled in a final volume of 700  $\mu$ l. RNase-free DNase (Boehringer-Mannheim) (70 U) was added and the mixture was incubated at 37 °C for 30 min to remove any contaminating DNA. The mixture was extracted with an equal volume of chloroform/isoamyl alcohol (24:1) and the RNA precipitated with three volumes of ethanol. Pellets were stored at -70 °C in 70 % ethanol and dissolved in water when required. RNA was quantified by UV-spectroscopy (if  $OD_{260} = 1.0$  then RNA = 30  $\mu$ g ml<sup>-1</sup>).

### 2.2.12 RNA dot blot analysis

RNA dot blots were performed on Hybond-N nylon membranes according to the manufacturer's instructions ("Blotting and hybridization protocols for Hybond membranes", Amersham). RNA from *E. coli* LK111 containing pUC18 and pNC181, *Rhodococcus* sp. ATCC 21145 and *Rhodococcus* sp. JL10 was heated to 90 °C for 2 min, cooled on ice and applied in spots containing 2  $\mu$ g, 5  $\mu$ g and 10  $\mu$ g of RNA onto Hybond-N (Amersham) nylon membrane which was sealed in a vacuum dot

blotting apparatus. The nylon membrane was then dried and fixed on a UV trans-illuminator as described by the manufacturer (Amersham).

Two probes, the whole plasmid pUC18 and the *Hind*III/*Bgl*II DNA fragment of the *Rhodococcus* insert from the pigment producing plasmid pKS1 (Fig. 3.1, Chpt3) were radioactively labelled using a nick translation kit following the manufacturers instructions (Amersham N.500). The probes were labelled with [ $\alpha$ - $^{32}$ P]dCTP with a specific activity of 3,000 Ci mmol $^{-1}$  (Amersham).

RNA isolated from *E. coli* LK111 (pUC18) was probed with labelled pUC18 DNA as a positive control and the *Rhodococcus Hind*III/*Bgl*II pigment-producing DNA fragment as a negative control. All other RNA dots on the Hybond-N membrane were probed with the labelled *Rhodococcus* DNA fragment from pKS1 and washed under stringent conditions following the manufacturers instructions (Amersham). Membranes were dried and exposed to X-ray film (Agfa, Curix) for 18-24 h.

## 2.3 RESULTS

### 2.3.1 Minimal media supplementation

*E. coli* strain 3.300 (pNC185) consistently and reliably produced copious amounts of pigments on LB Ap agar. *E. coli* strain 3.300 was used for metabolic studies because it does not have any amino acid growth requirements and does not produce *lac* repressor so that genes are constitutively expressed from the *lacOP* promoter of the pUC plasmids in *E. coli* 3.300 transformants.

The results of growing *E. coli* 3.300 (pNC185) on different media are summarised in Table 2.2. Colonies of *E. coli* 3.300 (pNC185) produced copious amounts of blue pigment on LB Ap agar plates whereas a control strain, *E. coli* 3.300 (pUC18), was unpigmented (Fig 2.1). Colonies of *E. coli* 3.300 (pNC185) on MM Ap lac plates, however, were unpigmented. LB agar is a complex growth medium containing tryptone and yeast extract. It was assumed, as an initial hypothesis, that at least one component of the complex LB broth and LB agar was a substrate for an enzyme encoded by the *Rhodococcus* DNA insert which was capable of transforming the substrate/s into one or more pigments.



**Fig. 2.1** Examples of  $\text{Pig}^+$  and  $\text{Pig}^-$  strains of *E. coli* transformants grown on LB Ap agar.

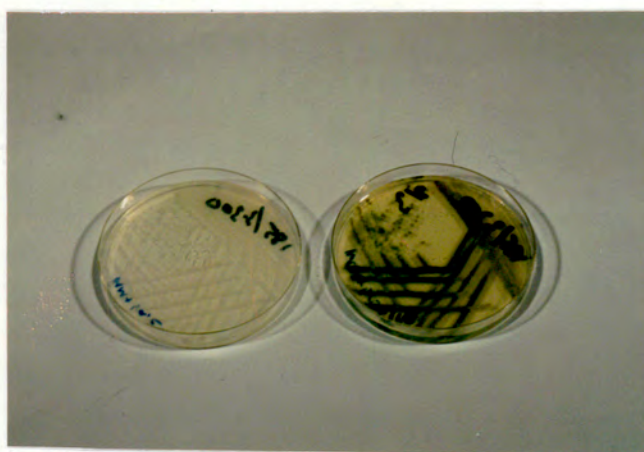
**Table 2.2** Pigmentation phenotypes of various bacterial strains on different agar growth media.

| STRAIN                            | GROWTH MEDIUM | PHENOTYPE        |
|-----------------------------------|---------------|------------------|
| <i>E. coli</i> 3.300 (pNC185)     | LB Ap         | $\text{Pig}^+$   |
| " " " "                           | MM lac Ap     | $\text{Pig}^-$   |
| " " " "                           | MM lac Ap Ind | $\text{Pig}^+$   |
| " " " "                           | MM lac Ap Trp | $\text{Pig}^+$   |
| <i>E. coli</i> MY252 (pNC185)     | LB Ap         | $\text{Pig}^+$   |
| <i>E. coli</i> MY252 (pNC185)     | LB Ap Ind     | $\text{Pig}^+$   |
| <i>E. coli</i> MY1393 (pNC185)    | LB Ap         | $\text{Pig}^-$   |
| <i>E. coli</i> MY1393 (pNC185)    | LB Ap Ind     | $\text{Pig}^\pm$ |
| <i>Rhodococcus</i> sp. ATCC 21145 | LB            | $\text{Pig}^-$   |
| <i>Rhodococcus</i> sp. ATCC 21145 | LB Ind        | $\text{Pig}^-$   |

As a first step to defining the growth medium requirements for pigment production, *E. coli* 3.300 (pNC185) was grown on MM lac Ap plates supplemented with tryptone ( $10 \text{ g l}^{-1}$ ) or yeast extract ( $5 \text{ g l}^{-1}$ ). Pigment was produced on MM lac Ap plates containing tryptone but not yeast extract which suggested that an amino acid was an essential component of the growth medium for pigment production since tryptone is particularly rich in peptides and amino acids.



The surfaces of MM lac Ap plates were spread with 100  $\mu$ l of 20 % solutions of one of the following amino acids: L-alanine, L-arginine, cysteine HCl, L-glutamic acid, L-glycine, L-histidine, L-leucine, L-methionine, D,L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine and D,L-valine and allowed to dry before streaking cells of *E. coli* 3.300 (pNC185) or *E. coli* 3.300 (pUC18) on them. It was observed that only *E. coli* 3.300 (pNC185) colonies on MM lac Ap plates which had been spread with tryptophan were pigmented (Fig. 2.2). All other plates supported growth of both *E. coli* strains but all colonies were white. This result suggested that the pigments may be based on indole, a breakdown product of tryptophan, which is known to have many pigmented derivatives. *E. coli* converts tryptophan to indole by the action of tryptophanase, an enzyme which is subject to catabolite repression (Botsford and DeMoss, 1971; Ward and Yudkin, 1976). Both *E. coli* 3.300 (pNC185) and 3.300 (pUC18) were grown on MM lac Ap plates containing 1 mM indole and both grew well but only *E. coli* 3.300 (pNC185) produced blue pigment which suggested that indole was a pigment precursor substance.



**Fig. 2.2** *E. coli* 3.300 (pNC185) on MM lac Ap (left) (Pig<sup>-</sup>) and MM lac Ap plus 1 mM tryptophan (right) (Pig<sup>+</sup>).

### 2.3.2 Transformation of a tryptophanase mutant strain of *E. coli*

The *E. coli* strain MY1393 ( $\Delta$  *trpAC9*, *tna* 103) is a mutant strain which, as well as having a growth requirement for tryptophan, does not produce tryptophanase (*tna*<sup>-</sup>) and therefore cannot produce indole from tryptophan, whereas the parent strain, MY252 ( $\Delta$  *trpAC9*, *tna*<sup>+</sup>) only has the tryptophan growth requirement and is wild



type with regard to tryptophanase activity. It was expected that if indole was required for pigment biosynthesis that the *tna*<sup>-</sup> mutant strain, MY1393, transformed with pNC185 would be non-pigmented whereas MY252 (*tna*<sup>+</sup>) (pNC185) would be pigmented. The phenotype of each strain was checked by attempting to grow the strains on MM agar containing tryptophan as the sole carbon and energy source, for which tryptophanase is required. Only strain MY252 (*tna*<sup>+</sup>) was able to grow under these conditions. Competent cells of both strains were prepared (Appendix A) and transformed with the plasmids pUC18 and pNC185. Transformants were selected on LB Ap agar and observed for pigment production (Table 2.2). Strain MY1393 (*tna*<sup>-</sup>) transformed with both pUC18 and pNC185 was non-pigmented as was MY252 (*tna*<sup>+</sup>) containing pUC18. Strain MY252 transformed with pNC185, however, produced blue pigmented colonies (Fig. 2.3). This result suggested that tryptophanase was important for pigment production and that indole was an essential precursor for pigment production. It was expected, therefore, that when *E. coli* MY1393 (pNC185) was grown on LB Ap agar plates containing 1 mM indole that pigment production would be restored. A low level of pigment production was obtained under these conditions but much less than obtained from *E. coli* MY252 (pNC185) (Fig. 2.3).



**Fig. 2.3** *E. coli* MY252 (pNC185) (right) and *E. coli* MY1393 (pNC185) (left) on LB Ap agar

### 2.3.3 *Rhodococcus* growth experiments

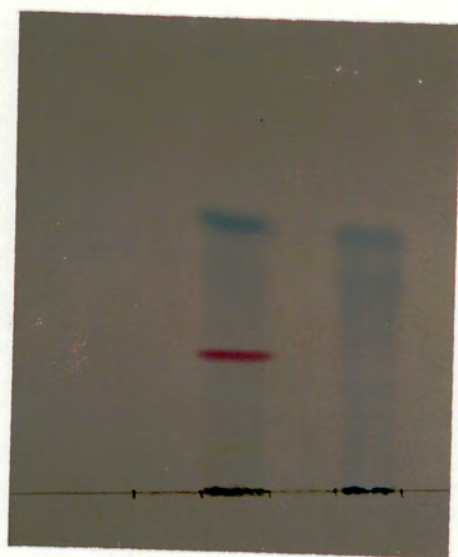
*Rhodococcus* sp. ATCC 21145 and *Rhodococcus* sp. JL10 grown on LB agar, LB agar plates containing 1 mM indole or in LB broth containing 1 mM indole did not produce blue pigment.

### 2.3.4 *In vitro* assay for pigment production

No pigment production was detected by cell-free extracts of pigment-producing *E. coli* transformants under the conditions described.

### 2.3.5 Thin layer chromatography

Samples of pigmented bacterial chloroform extracts were analysed by TLC and compared with a sample of genuine indigo. The bacterial pigment separated into a faster moving blue component and a slower pink component (Fig. 2.4).



**Fig. 2.4** TLC of bacterial pigments (left) and indigo (right).

The indigo sample contained a predominant blue component with the same mobility as the blue component of the bacterial pigment extract ( $R_F = 0.21$ ), and the indigo sample also exhibited a very faint pink component with the same mobility as the pink bacterial pigment ( $R_F = 0.07$ ). It was observed that in freshly prepared



bacterial pigment extracts the blue component was predominant but in older extracts the pink fraction increased in concentration relative to the blue fraction.

### 2.3.6 Pigment absorption spectra

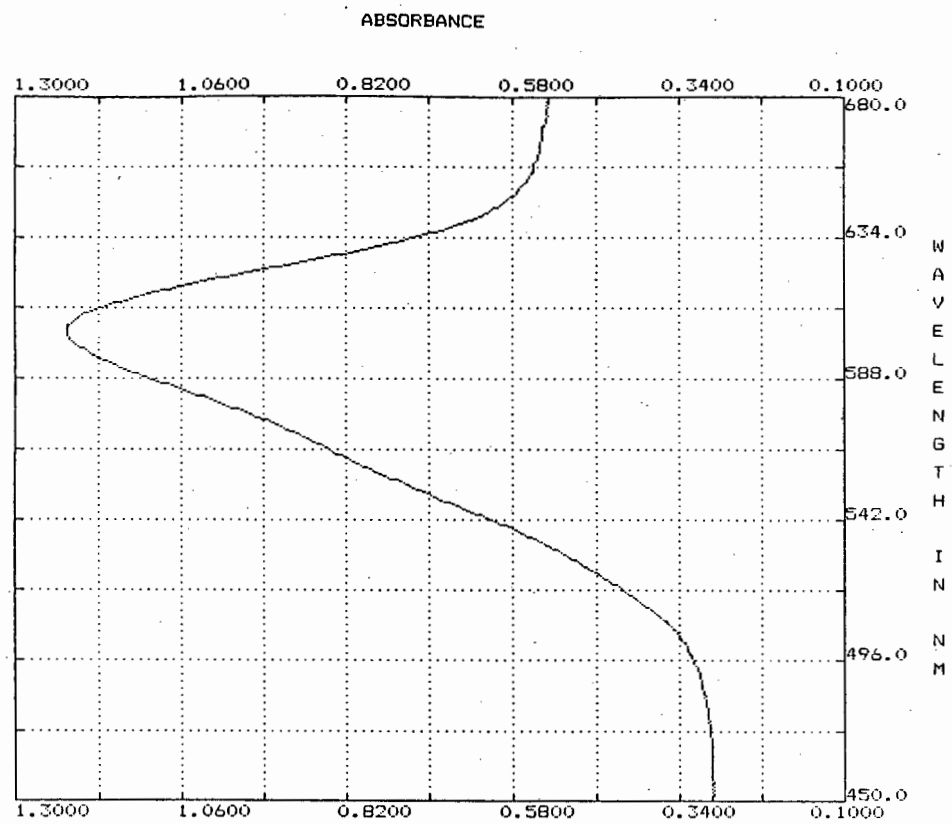
The absorption spectra of pigment solutions in chloroform purified by preparative TLC were determined (Fig. 2.5). The blue pigment had an absorption peak at 602 nm whereas the pink pigment absorbed light most strongly at 566 nm. The absorption spectrum of a chloroform solution of indigo was similar to that of the blue bacterial pigment and also had an absorption peak at 602 nm (Fig. 2.5). The absorption spectrum scan of a freshly extracted, pigmented chloroform solution which had not been separated by TLC, had a major absorption peak at 602 nm with a shoulder in the peak at approximately 560 nm which was presumably caused by the combination of the pink and blue pigments in the same solution (Fig. 2.5). The unpurified pigment extract was allowed to stand at room temperature overnight for about 16 h after which time the pigment extract had taken on more of a purplish hue. An absorption scan showed that the predominant absorption peak had shifted to 566 nm and a shoulder was apparent at approximately 600 nm (Fig. 2.5). This observation suggested that the blue pigment was slowly being converted into the pink pigment on standing in the crude bacterial chloroform extract.

The TLC-purified pink and blue pigments appeared to be stable in chloroform. A sample of genuine indigo was dissolved in a non-pigmented chloroform extract of *E. coli* LK111 (pUC18) and left at room temperature for 16 h, however, no change was seen in the absorption spectrum of the indigo solution.

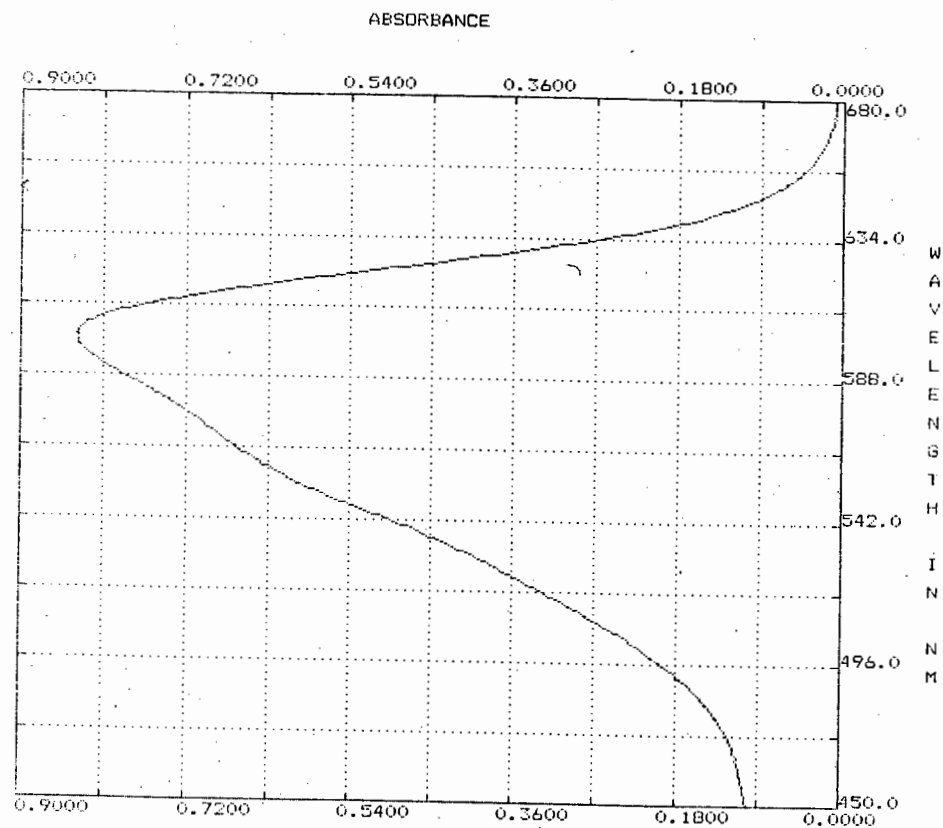
These results suggested that the blue bacterial pigment may be indigo or an indigo-related compound and that the pink pigment was chemically related to the blue pigment.

### 2.3.7 Chemical properties of the pigments

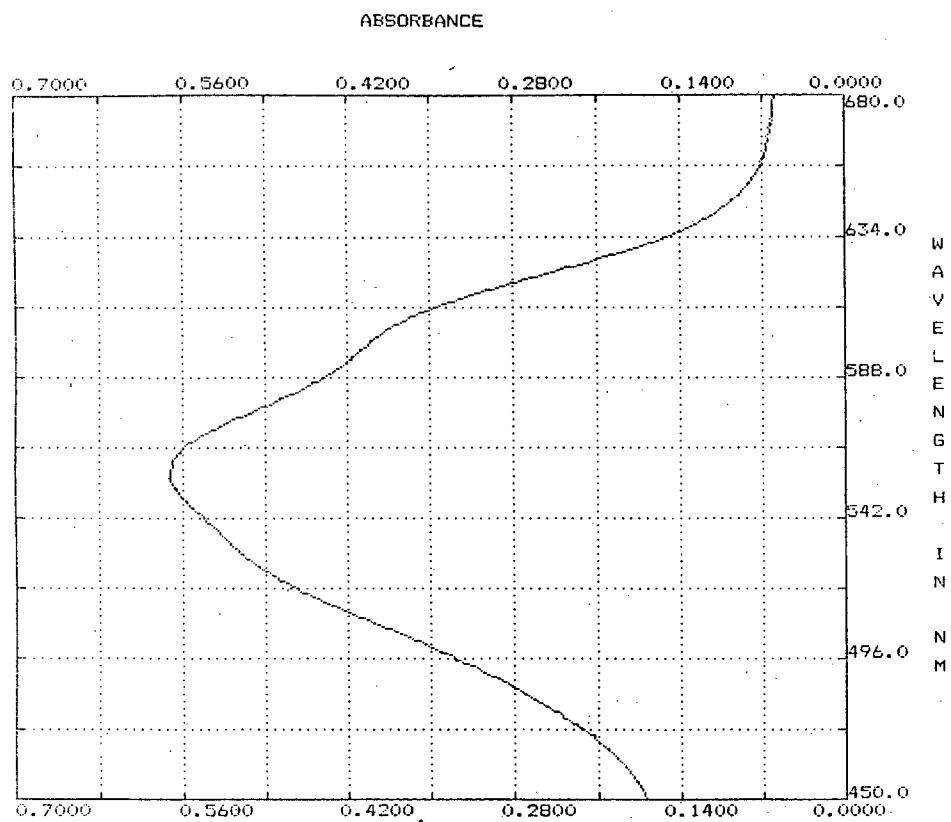
Freshly prepared, pigmented chloroform extracts (which were not TLC separated) were evaporated and redissolved in a range of solvents. The pigment was insoluble in water, ethanol and methanol.



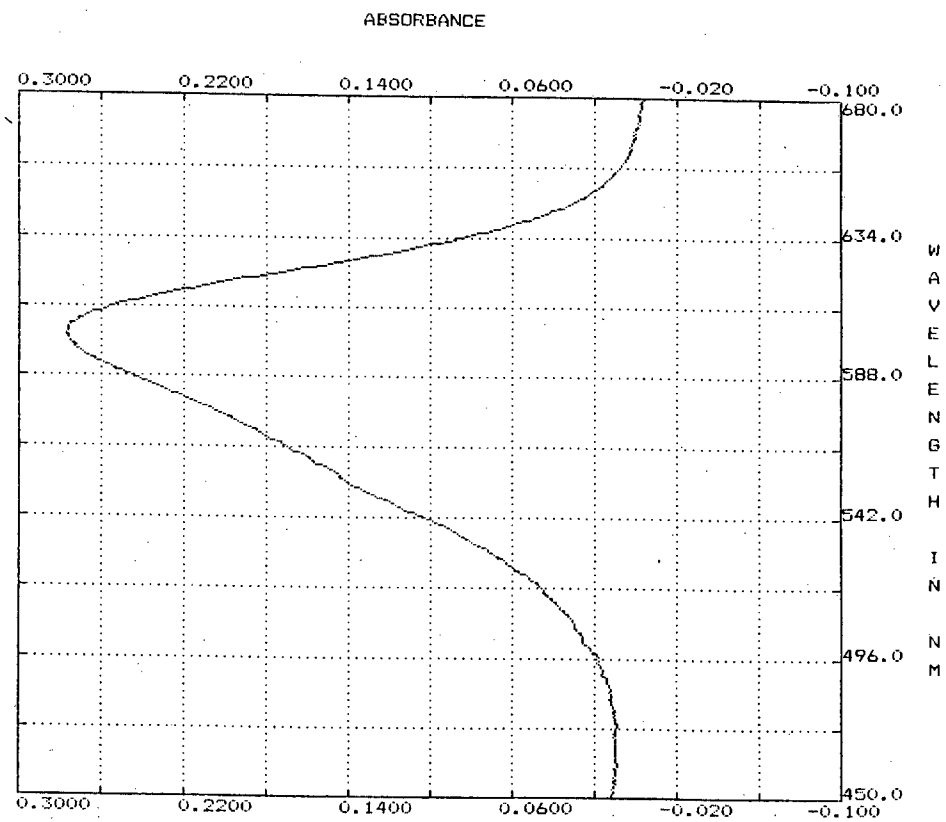
1



2

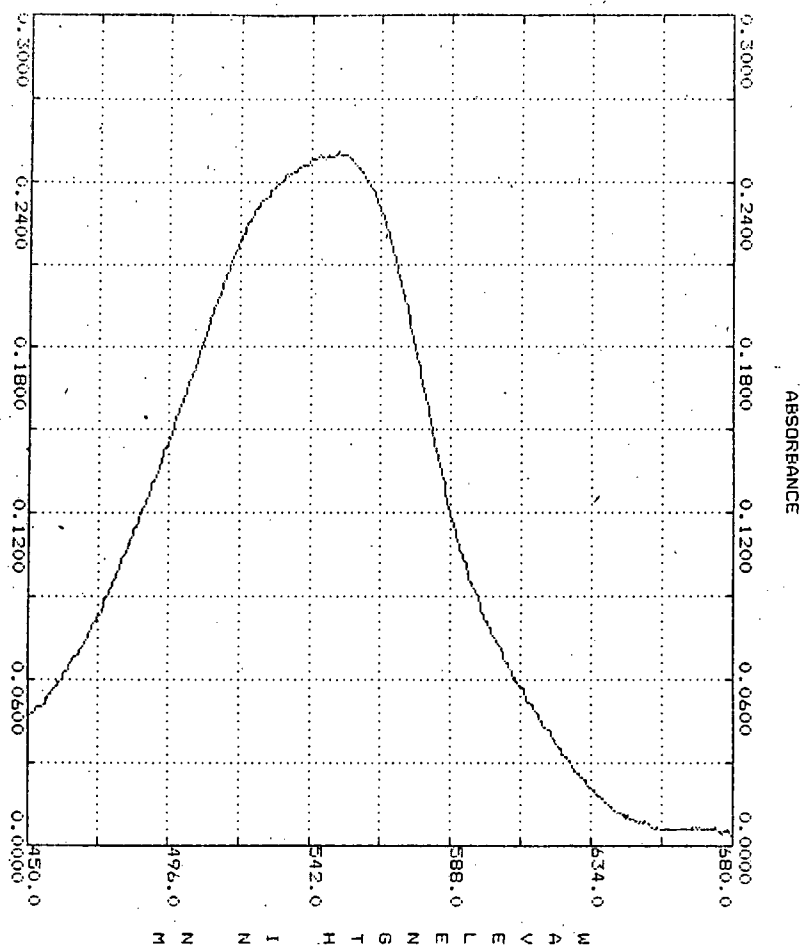


3



4

5



**Fig 2.5** Absorption spectra of purified and unpurified pigments dissolved in chloroform. 1) indigo in a chloroform extract of *E. coli* 3.300 (pUC18), (Pig<sup>-</sup>); 2) fresh, pigmented chloroform extract of *E. coli* 3.300 (pNC185) (Pig<sup>+</sup>); 3) 2) after aging; 4) TLC-purified blue pigment ; 5) TLC-purified pink pigment.

In other organic solvents, the pigment was insoluble in octane, slightly soluble with a reddish colour in both xylene and toluene and very soluble in acetone (a reddish blue solution), chloroform (a purple solution), 1,2-dichloroethane (a red solution), ethylmethylketone (a reddish blue solution) and glacial acetic acid (a reddish blue solution). The best solvent was the highly polar organic solvent dimethyl sulphoxide (DMSO) which dissolved the pigment to give a deep azure-blue solution. Indigo had similar solubility properties although the colours of the bacterial pigment solutions were redder in solvents such as acetone possibly due to the greater proportion of pink pigment in the unpurified bacterial pigment extract than in commercial indigo.

Adjustment of the pH had no effect on the colour of purified blue and pink pigments and no other direct chemical means of inter-converting the two pigments was found.

### 2.3.8 NMR spectroscopy

It was possible to conclude from preliminary  $^1\text{H}$  NMR spectroscopy of TLC-purified bacterial pigments that the blue pigment had some features consistent with an indole type pigment although it was not possible to say that it was definitely indigo (Appendix C, personal communication, Professor K. Koch, Dept. of Chemistry, University of Cape Town). A large amount of aliphatic material was consistently associated with the aromatic chromophore in the case of both blue and pink components. It was not possible to establish whether the aliphatic material was simply contaminating bacterial matter (e.g. cell wall component(s)) not removed by the TLC pigment purification procedure, or if the association was due to chemical bonding or the formation of a stable pigmented conjugate. The pink pigment was analysed by both  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy and both analyses suggested that the structure of the pink compound consisted of two indole-type structures linked in a manner which resulted in two separate spin systems being evident and was compatible with the structure of indirubin, an asymmetrical, structural isomer of indigo. An attempt to obtain a  $^1\text{H}$  NMR spectrum of authentic indigo failed, presumably because some unpaired electron density severely broadened all resonances.

### 2.3.9 RNA dot blot analysis

The aim of this experiment was to establish whether or not the *Rhodococcus* gene contained within the pigment-producing DNA insert in *E. coli* plasmids such as pNC185 was being expressed in the original *Rhodococcus* sp. since it did not produce the blue pigment. The positive control probe, the plasmid pUC18 hybridized strongly with RNA from *E. coli* LK111 (pUC18) (Fig. 2.6). The negative control, the *Hind*III/*Bgl*II fragment from the *Rhodococcus* DNA insert in pKS1, failed to hybridize with the same RNA. The RNA from *E. coli* LK111 (pNC181) hybridized strongly with the *Hind*III/*Bgl*II *Rhodococcus* DNA fragment, as expected. RNA from *Rhodococcus* sp. ATCC 21145 hybridized weakly with the *Hind*III/*Bgl*II fragment

whereas RNA from another rhodococcus strain, *Rhodococcus* sp. JL10, showed no hybridization signal at all with the same probe under stringent hybridization and washing conditions (Fig. 2.6). These results suggested that the same DNA insert which was transcribed and caused pigmentation in recombinant *E. coli* was also transcribed in *Rhodococcus* sp. ATCC 21145 but did not allow pigmentation in this bacterium.

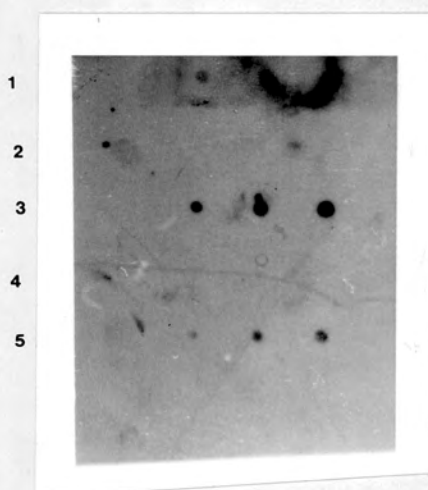


Fig. 2.6 RNA dot blot. 2  $\mu$ g, 5  $\mu$ g and 10  $\mu$ g of each RNA sample were loaded onto the membrane and probed with the *Hind*III/*Bgl*II fragment of the *Rhodococcus* pigment-producing insert in pKS1 unless stated otherwise. 1) *E. coli* LK111 (pUC18) probed with pUC18 DNA; 2) *E. coli* LK111 (pUC18); 3) *E. coli* LK111 (pNC181); 4) *R. corallinus* JL-10; 5) *Rhodococcus* sp. ATCC 21145.



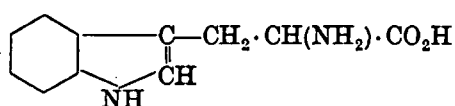
## 2.4 DISCUSSION

It is proposed that the blue component of a pigment produced by *E. coli* containing a cloned *Rhodococcus* gene is indigo and that the pink component is indirubin, an indigo isomer. The biosynthetic pathway probably involves the enzymatic conversion of tryptophan to indole by tryptophanase of *E. coli*, followed by the oxidation of indole by the product of the cloned *Rhodococcus* gene in *E. coli*. Evidence for this theory consisted of the following observations :

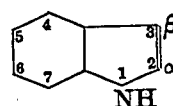
- 1) Tryptophan and indole were shown to be precursor substrates for pigment production by *E. coli* containing pigment-producing plasmids on minimal medium.
- 2) A tryptophanase mutant *E. coli* strain, MY1393, which was unable to produce indole from tryptophan, was non-pigmented when transformed with the Pig<sup>+</sup> plasmid pNC185.
- 3) The blue pigment had an identical  $R_F$  value by TLC to indigo and a pink pigment faintly observed in the TLC-separated indigo solution which was probably indirubin had the same mobility as the pink bacterial pigment.
- 4) The purified blue bacterial pigment had an identical absorption spectrum to indigo at visible wavelengths.
- 5) The solubility characteristics of the unpurified bacterial pigment and indigo were very similar in a wide range of different solvents.
- 6) Proton NMR spectroscopy suggested that both the blue and pink pigments were based on indole, and resonance patterns were compatible with the structures of indigo and indirubin for the blue and pink bacterial pigments, respectively.

The identical mobility by TLC of the blue and pink bacterial pigments with the major blue and faint pink components of indigo, suggested that they were identical compounds, indigo (blue) and indirubin (pink), and that the aliphatic material detected by NMR spectroscopy was due to contamination not removed by the TLC

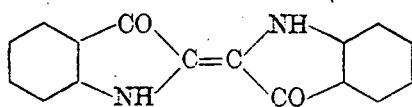
procedure rather than chemical bonding or conjugation. The difference in relative amounts of the pink and blue components in the commercial indigo and the bacterial pigment extract may be due to differences in pigment isolation procedures or, alternatively, the aliphatic material may indeed be associated in a more permanent way with the pigments which stabilises the pigment compound as the pink form, indirubin.



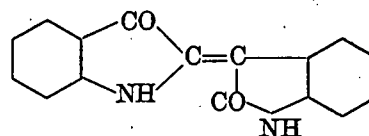
Tryptophan



Indole



Indigo



Indirubin

Fig. 2.7 Molecular structure of indigo and related compounds.

Tryptophan and indole were shown to be precursor substrates for pigment production by *E. coli* containing pigment-producing plasmids which suggested that indigo was being produced by the same pathway suggested by Ensley *et al.* (1983) (Fig. 2.8) in which tryptophan was converted to indole by the constitutively expressed tryptophanase gene of *E. coli* and indole was then oxidized by the cloned gene product, naphthalene dioxygenase, to indoxyl which underwent autoxidation in air to indigo. The conversion of tryptophan to indole by tryptophanase was demonstrated to be a step in the pathway to pigment production by transformation

of pNC185 (Pig<sup>+</sup>) into a tryptophanase mutant *E. coli* strain which failed to produce pigments.

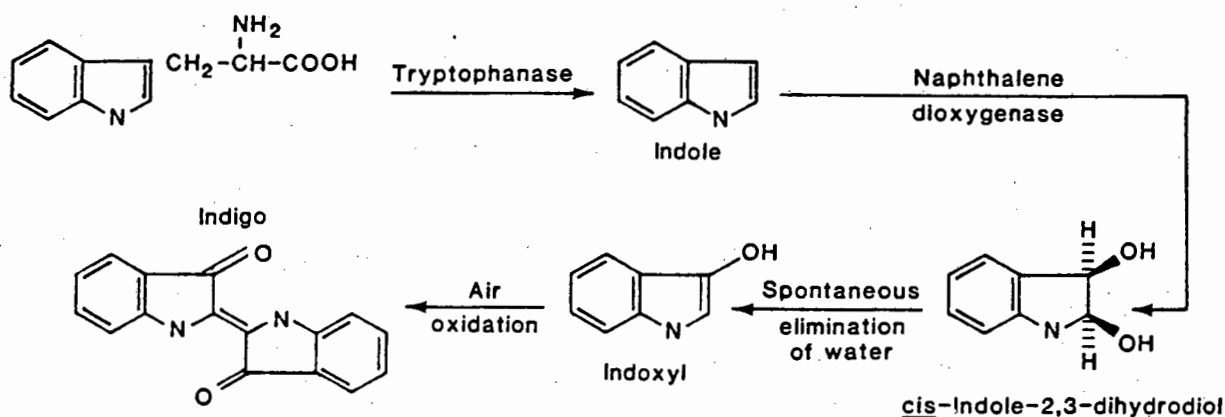


Fig. 2.8 The pathway of bacterial indigo biosynthesis from tryptophan (Ensley *et al.*, 1983).

The formation of indigo by *Rhodococcus* sp. ATCC 21145 was not demonstrated under the defined conditions although the RNA dot blot analysis suggested that the pigment-forming gene was being transcribed in *Rhodococcus*. This may be due to the low level of activity of the gene at the single-copy level or possibly due to the rapid transformation of the pigment into another product by *Rhodococcus* enzymes preventing accumulation of the pigment.

Indigo production by *E. coli* containing cloned genes from *Pseudomonas* spp. has been used as an indicator for expressed aromatic dioxygenase genes from a wide variety of pathways (Eaton and Timmis, 1986; Ensley *et al.*, 1983; Irie *et al.*, 1987; Tan and Mason, 1990; Zylstra *et al.*, 1988) although none of these authors reported the observation of pink as well as blue components in the bacterial pigment. This

suggests that the enzyme encoded by the pigment-producing *Rhodococcus* gene was an aromatic dioxygenase. These enzymes are important components of pathways in the degradation of aromatic, xenobiotic compounds and in the biosynthesis of useful products, such as salicylic acid from naphthalene (Ensley *et al.*, 1983). *Rhodococcus* spp. display great metabolic diversity and are capable of degrading a wide range of aromatic compounds (Chapter 1) although the exploitation of their full potential in this area has hardly begun. The test for indigo production may be a useful tool in the search for other *Rhodococcus* genes encoding aromatic dioxygenases, or even whole operons. The function and substrate range of the proposed aromatic dioxygenase from *Rhodococcus* sp. ATCC 21145, which causes indigo production in *E. coli*, has not yet been established although this *Rhodococcus* strain was described in a patent (Raymond, 1971) for the production of hydroxyphenylketobutyric acids by the microbial oxidation of naphthalene.

## CHAPTER THREE

## NUCLEOTIDE SEQUENCE OF THE PIGMENT GENE

|  |    |
|--|----|
| 3.1 INTRODUCTION   | 46 |
| 3.2 METHODS AND MATERIALS  | 49 |
| 3.2.1 Bacterial strains, plasmids and media  | 49 |
| 3.2.2 Preparation of plasmid DNA   | 50 |
| 3.2.3 Subcloning and restriction endonuclease digestions                           | 50 |
| 3.2.4 DNA sequencing   | 51 |
| 3.2.5 <i>In vitro</i> transcription translation                                    | 52 |
| 3.3 RESULTS  | 52 |
| 3.3.1 Restriction fragment deletions and subclonings                               | 55 |
| 3.3.2 Nucleotide sequencing  | 55 |
| 3.3.3 Codon usage and the ORF  | 57 |
| 3.3.4 Upstream DNA sequence analysis with <i>lac</i> fusions                       | 60 |
| 3.3.5 Deduced amino acid sequence and <i>in vitro</i><br>transcription-translation | 61 |
| 3.4 DISCUSSION   | 62 |

## CHAPTER THREE

### NUCLEOTIDE SEQUENCE OF THE PIGMENT GENE

**SUMMARY:** The 2.1 kbp DNA fragment from *Rhodococcus* sp. ATCC 21145 which produced the blue and pink pigments in *E. coli* when cloned downstream of a strong promoter, was sequenced and contained an apparent open reading frame (ORF) of 1,653 bp. The coding sequence for the pigment-producing enzyme was shown to consist of an actual ORF of 1,161 bp, preceded by a putative ribosome binding site which had a potential coding capacity for a single protein of  $M_r$  42,560. Deletion analysis, codon usage analysis and *in vitro* transcription-translation analysis experiments supported the hypothesis that the 1,161 bp coding sequence produced a single enzyme responsible for pigment production in *E. coli*. It was shown that the upstream sequence of the *Rhodococcus* DNA did not contain *E. coli*-type promoter sequences and was not capable of regulating pigment gene expression in *E. coli*. Sequence comparisons with nucleotide and protein data bases did not reveal any similarities with the *Rhodococcus* pigment gene.

### 3.1 INTRODUCTION

The plasmid pNIL260, which contained a 2.8 kbp *Rhodococcus* DNA insert, conferred the Pig<sup>+</sup> phenotype on *E. coli* transformants (Hill *et al.*, 1989). The aims of the studies reported in this chapter were to determine the genetics of pigment production in *E. coli* (pNIL260). The initial approach was to determine and analyse the nucleotide sequence of the smallest piece of cloned *Rhodococcus* DNA required for pigment production in *E. coli*. It was hoped that similarity of the nucleotide sequence and deduced amino acid sequence to proteins of known function in computer data bases may suggest possible pathways for pigment production which could then be tested. Knowledge of the nucleotide sequence also would facilitate further genetic manipulations of the *Rhodococcus* pigment-producing DNA insert. The determination of the *Rhodococcus* pigment gene sequence and the analysis of the features contained within the sequence also were important because *Rhodococcus* spp. have been poorly studied at the genetic level (Chapter 1.2). It is known that the DNA of *Rhodococcus* spp. has a high G+C content in the range 59-69 mol % (Goodfellow and Minnikin, 1977) although the closely related *Streptomyces* have an even higher G+C content of 73 mol % (Enquist and Bradley, 1971). Among bacterial species, the mean G+C content of genomic DNA varies from approximately 25-74 mol %. The diversification of G+C content probably played a major role in evolution of the genome (Muto and Osawa, 1987). The *Streptomyces* spp. appear to have accommodated the high G+C content by the use of codons with a G or C in the third position which is often allowed by the degeneracy of the genetic code. The first codon position also allows the preferential use of G or C but to a lesser extent than the third position while the second codon position does not allow any substitutions. This results in a predictable G+C composition at each of the three codon positions which has been used to identify protein coding sequences and the direction of transcription (Bibb *et al.*, 1984). It is not clear whether the codon usage pattern of *Streptomyces* evolved as a consequence of the high G+C content of the DNA or whether the base composition reflects an adaptation to a particular pool of tRNA species. Codon usage in *E. coli* has been shown to be a mechanism for regulating gene expression (Ernst, 1988) and in most organisms there are strong

preferences amongst synonymous codons, particularly in highly expressed genes (Ikemura, 1985). In addition the sequences flanking a codon have been shown to affect gene expression (Gouy, 1987; Shpaer, 1986; Yarus and Folley, 1985) through both structural and translational effects. Context is important for determining start codons (Stormo, 1986) and the efficiency of suppression of nonsense codons (Bossi, 1983; Bossi and Roth, 1980; Miller and Albertini, 1983), and the context of the stop codon UGA can cause it to be read as selenocysteine, the twenty first amino acid (Zinoni *et al.*, 1987).

Pigment production in *E. coli* by a cloned fragment of *Rhodococcus* DNA provided a potentially simple system for the assay of a putative *E. coli*-functional *Rhodococcus* promoter (Hill *et al.*, 1989). *Rhodococcus* promoters have not been studied before but their expression in *E. coli* was considered unlikely because of the close relationship between *Rhodococcus* spp. and *Streptomyces* spp. *E. coli* transcriptional units possess characteristic DNA sequences situated 5' to the coding sequence. These are the -10 region, or Pribnow box, which has a consensus sequence of 5' -TATAAT- 3', and the -35 region which has a consensus sequence of 5' -TTGACA- 3'. The -35 region is considered to be the recognition site for the binding to DNA of RNA polymerase as a first step in transcription while the -10 region signals to the RNA polymerase which strand of DNA is to be transcribed and in which direction. *Streptomyces* genes are usually not expressed from their own promoters in *E. coli* because they lack the upstream -10 and -35 recognition sequences. *Streptomyces* have different classes of promoters which are recognised by different forms of RNA polymerase (Westpheling, 1985) although a form of RNA polymerase was found in *Streptomyces* which was capable of recognizing typical *E. coli*-type promoters (Bibb *et al.*, 1985) and a few cases have been described of *Streptomyces* genes being expressed in *E. coli* (Buttner and Brown, 1987; Deng *et al.*, 1986; Riverolezcano *et al.*, 1990). The number of *Streptomyces* promoters which can be expressed in *E. coli*, however, is estimated at less than 10 % (Jaurin and Cohen, 1985).

Evidence from *Micrococcus luteus* suggested that a high G+C content did not exclude the possibility of *E. coli*-type promoters. The nucleotide sequence of the 5S ribosomal RNA from *R. erythropolis* has been published and used for phylogenetic analysis. It was shown that *R. erythropolis* is more closely related to *M. luteus* than to *Mycobacterium tuberculosis* (Park *et al.*, 1987). The G+C content of *M. luteus* is 74 %



yet promoter sequences at the -10 and -35 positions of two *M. luteus* genes were shown to be similar to those of *E. coli* and an *in vitro* transcription assay showed that *E. coli* RNA polymerase was able to transcribe correctly from both of these *Micrococcus* promoters (Nakayama *et al.*, 1989).

Among other actinomycetes, the expression of mycobacterial DNA in *E. coli* was studied and some evidence was obtained for the expression of mycobacterial promoters in *E. coli* (Labidi *et al.*, 1985; Kieser *et al.*, 1986).

## 3.2 METHODS AND MATERIALS

### 3.2.1 Bacterial strains, plasmids and media

The *E. coli* strain LK111 was used for all experiments (K514 derivative; *lacI*, *lacZ*,  $\Delta$  M15, *lacY*+) (Zabeau and Stanley, 1982). The plasmids used and constructed are listed in Table 3.1. The phenotype of the plasmids was designated  $\text{Pig}^+$  or  $\text{Pig}^-$  depending on whether they could or could not produce pigment in *E. coli*.

Table 3.1 Plasmids used in this study.

| PLASMID        | PHENOTYPE                                  | SOURCE                         |
|----------------|--|--------------------------------|
| "Bluescript"KS | $\text{Ap}^r \text{ LacZ}^+$               | Stratagene, San Diego, USA     |
| pBSLac1        | $\text{Tet}^r \text{ LacZ}^-$              | This study                     |
| pHSLac1        | $\text{Tet}^r \text{ LacZ}^+$              | This study                     |
| pKS1           | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pKS3           | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pMC1871        | $\text{Tet}^r \text{ LacZ}^-$              | Casadaban <i>et al.</i> (1983) |
| pNC181         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pNC184         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^-$ | This study                     |
| pNC185         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pNC186         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^-$ | This study                     |
| pNC191         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pNC192         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pNC194         | $\text{Ap}^r \text{ LacZ}^- \text{ Pig}^+$ | This study                     |
| pUC18          | $\text{Ap}^r \text{ LacZ}^+$               | Vieira & Messing (1982)        |
| pUC19          | $\text{Ap}^r \text{ LacZ}^+$               | Vieira & Messing (1982)        |

*E. coli* LK111 containing pUC or "Bluescript"-derived plasmids was maintained on LB agar (Appendix B) containing ampicillin (Ap) at a concentration of  $100 \mu\text{g ml}^{-1}$ . Transformants containing derivatives of pMC1871 were selected on LB agar containing tetracycline (Tc) at a concentration of  $12.5 \mu\text{g ml}^{-1}$ , X-Gal ( $40 \mu\text{g ml}^{-1}$ ) and IPTG ( $12 \mu\text{g ml}^{-1}$ ), or on lactose MacConkey agar (Appendix B) containing Tc ( $12.5 \mu\text{g ml}^{-1}$ ).

### 3.2.2 Preparation of plasmid DNA

Plasmid DNA was prepared from *E. coli* by the alkaline lysis method of Ish-Horowicz and Burke (1981). Large scale DNA preparations were purified by CsCl equilibrium gradient centrifugation (Maniatis *et al.*, 1982). Preparations for DNA sequencing were treated with RNase A (Maniatis *et al.*, 1982) followed by phenol/chloroform extraction (Maniatis *et al.*, 1982) and ethanol precipitation (Maniatis *et al.*, 1982). Plasmid DNA pellets were dissolved in TE buffer (Appendix B) and the concentration of DNA determined by UV-spectroscopy (Maniatis *et al.*, 1982).

### 3.2.3 Subcloning and restriction endonuclease digestions

Restriction endonucleases (Boehringer-Mannheim, Anglian Biotechnology or BRL) were used as described by Maniatis *et al.* (1982) (Appendix A). Digested DNA fragments were separated in agarose gels (0.8 % w/v) containing EtdBr (0.5 g ml<sup>-1</sup>) by electrophoresis (2-5 V cm<sup>-1</sup>) in TBE buffer (Appendix B) and analysed on a UV transilluminator (Maniatis *et al.*, 1982). Gel-purification of DNA fragments (Appendix A) was performed by excision of EtdBr-stained bands in low-melting point agarose gels (Seaplaque) after electrophoresis in TAE buffer (Appendix B). The DNA was separated from the agarose with "Geneclean" (Bio 101, California, USA) according to the manufacturer's instructions. Klenow fragment of DNA polymerase I (Boehringer-Mannheim) or T4 DNA polymerase (Boehringer-Mannheim) were used to generate blunt-ended restriction fragments (Maniatis *et al.*, 1982). Ligations were performed with 1 U of T4 DNA ligase (Boehringer-Mannheim) in a mixture containing 2 µl of 10 x ligation buffer (Appendix B), vector and insert DNA and water up to a final volume of 20 µl. Self ligation of the vector plasmids was prevented by treatment with calf intestinal alkaline phosphatase (Boehringer-Mannheim) (Tabor, 1987). The DNA products of the ligation reaction were used to transform *E. coli* competent cells (Appendix A) and transformants selected on LB agar containing the appropriate antibiotic.

### 3.2.4 DNA sequencing

DNA was prepared for sequencing by the construction of ordered deletions of the pigment gene in both directions in the plasmid pKS3 using exonuclease III (Boehringer-Mannheim) (Henikoff, 1984). Approximately 6 µg of pKS3 plasmid DNA were digested at opposite ends of the pigment gene with a restriction endonuclease which cleaved the DNA in the cloning cassette generating a 3' overhang at the upstream end of the insert. This digested DNA was then cleaved again in the cloning cassette at a restriction site closer to the insert than the first cleaved site with a restriction endonuclease which generates a 5' overhang. Plasmid pKS3 was digested in the "Bluescript" multiple cloning cassette at the *Cla*I and *Apa*I restriction sites at one end while the restriction sites for *Xba*I and *Sac*I were used at the other end. The restriction enzyme digested DNA was extracted with phenol/chloroform (Maniatis *et al.*, 1982), precipitated with ethanol (Maniatis *et al.*, 1982) and resuspended in 60 µl of exonuclease III buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl<sub>2</sub>, 10 µg ml<sup>-1</sup> tRNA). Subsequently, the smaller, upstream *Bgl*III/*Sph*I fragment of pNC192 was subcloned into pUC19 to create the plasmid pBS1 from which exonuclease III shortenings were prepared to obtain more sequence in the 5' non-coding region of the *Rhodococcus* DNA insert.

Ten microlitres of pKS3 DNA (1 µg) in exonuclease III buffer were removed to the "Time = 0" microfuge tube containing 40 µl of 5 x Mung bean nuclease buffer (150 mM sodium acetate, pH 5.0, 250 mM NaCl, 5 mM ZnCl<sub>2</sub> and 25 % glycerol) and 150 µl of water on ice. The rest of the DNA was equilibrated to 37 °C and 600 U of exonuclease III added. After a 30 s lag period counting was started and every 30 s after that 10 µl samples of the exonuclease III reaction mixture were removed to microfuge tubes on ice containing diluted 5 x Mung bean nuclease buffer. When all samples were collected exonuclease III was heat inactivated at 70 °C for 10 min. Fifteen units of Mung bean nuclease in 1 x Mung bean nuclease dilution buffer (10 mM sodium acetate, pH 5.0, 0.1 mM zinc acetate, 1 mM cysteine, 0.1 % Triton X-100, 50 % glycerol) were added to each tube and incubated at 30 °C for 30 min to remove the single stranded DNA remaining after exonuclease III digestion. The reaction mixtures were extracted with phenol/chloroform (Maniatis *et al.*, 1982), the DNA precipitated with ethanol (Maniatis *et al.* 1982) and resuspended in 15 µl of TE buffer (Appendix B). The DNA was self-ligated in a final volume of 20 µl as

described previously and 10  $\mu$ l of the ligation mixture was used to transform *E. coli* LK111 competent cells (Appendix A). Transformants were selected on LB Ap agar plates.

DNA sequencing was done by the dideoxy chain-termination method of Sanger *et al.*, (1977) with a "Sequenase" kit (US Biochemicals Corp.) according to the manufacturers instructions. The DNA was radiolabelled with [ $^{35}$ S]dATP [ > 1000 Ci (37 kBq) mmol $^{-1}$ ] (Amersham, UK). The nucleotide sequences were analysed with a VAX main frame computer using the GCG sequence analysis package (Devereux *et al.*, (1984) version 6.2, August, 1990) as well as an IBM XT computer, using the DNA Tools and Genepro (version 3.1) programs. Deduced amino acid sequences were analysed and compared by using the data bases of GenBank (release 65.0, September, 1990), EMBL (release 24.0, August, 1990), NBRF-nucleic (release 36.0, April, 1990), NBRF-Protein (release 25.0, June, 1990) and Swiss-Protein (release 15.0, August, 1990).

### 3.2.5 *In vitro* transcription translation

A prokaryotic DNA-directed *in vitro* transcription-translation kit (Amersham, UK) was used according to the manufacturers instructions for the *in vitro* transcription and translation of plasmid DNA. L-[ $^{35}$ S] Methionine [ > 1000 Ci (37 kBq) mmol $^{-1}$ ] (Amersham, UK) was used as the radiolabelled amino acid. The labelled proteins were separated by electrophoresis in a 15 % (w/v) polyacrylamide gel containing 0.1 % SDS by the method of Laemmli (1970). High molecular weight standards ( $M_r$  18,500-330,000) (Pharmacia) were used to estimate the size of labelled proteins.

## 3.3 RESULTS

### 3.3.1 Restriction fragment deletions and subclonings

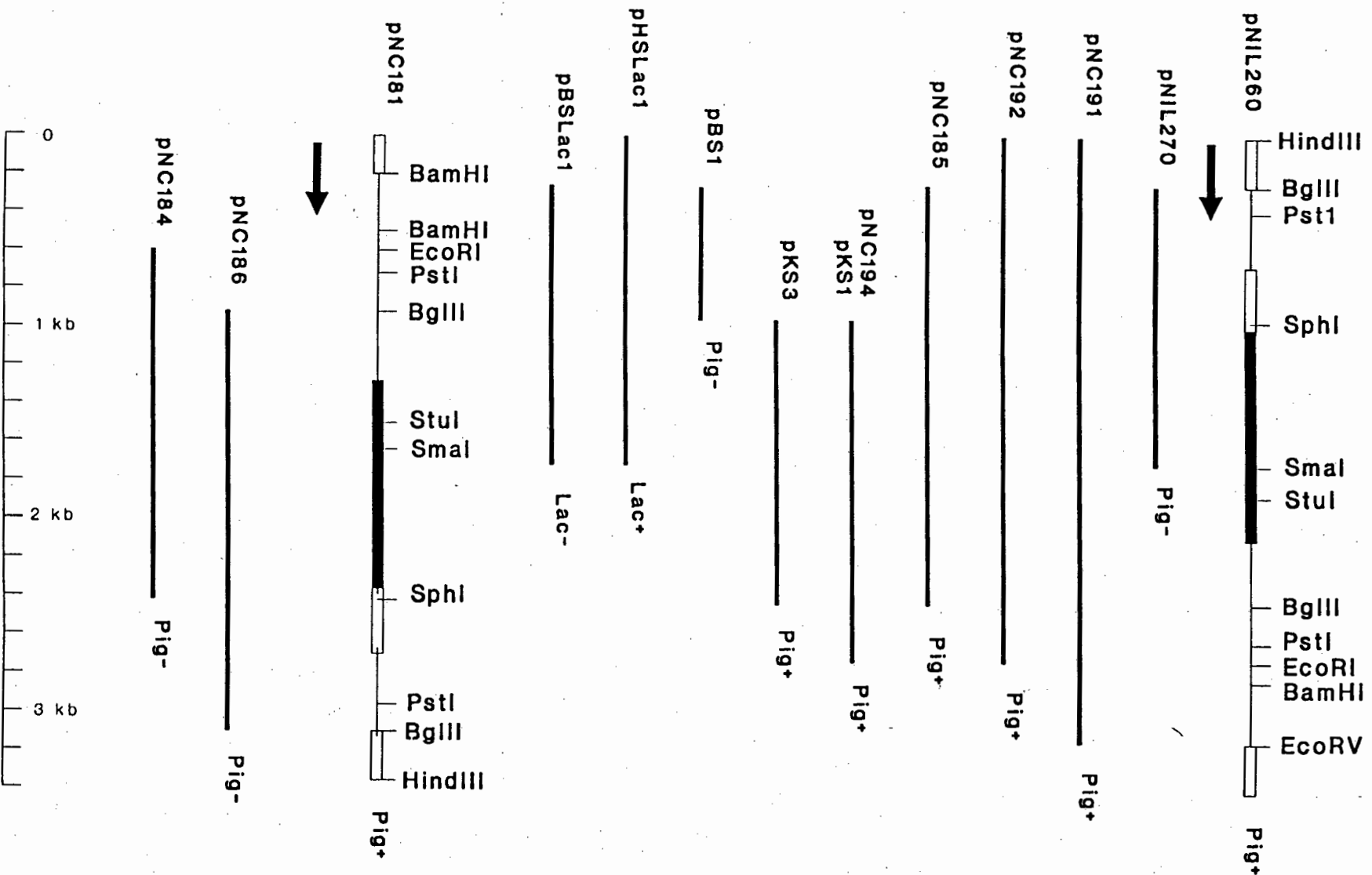
Restriction maps and plasmids used in the analysis and sequencing of the pigment gene are summarised in Fig. 3.1. The original pigment-producing plasmid from the gene bank, pNIL200 (15.1 kbp insert), was shortened to produce the Pig $^+$  plasmid pNIL260 (2.8 kbp insert) (Hill *et al.*, 1989). Deletion of the *Sma*I/*Eco*RV fragment of pNIL260 produced the Pig $^-$  plasmid pNIL270 which suggested that at least a part of

the reading frame for the pigment gene had been deleted. The 3.1 kbp *Hind*III/*Eco*RV fragment of pNIL260 was subcloned into pUC18 and pUC19 to produce the Pig<sup>+</sup> plasmids pNC181 and pNC191, i.e. the pigment-producing DNA insert was subcloned in both orientations in relation to the *lacOP* promoter of the pUC plasmids and was Pig<sup>+</sup> in both orientations suggesting that a *Rhodococcus* promoter was functioning in *E. coli*.

The plasmid pNC192 was generated by deletion of a restriction fragment of about 200 bp between the *Eco*RI restriction sites in the *Rhodococcus* DNA insert of pNC191 (Fig. 3.1) and in the multiple cloning cassette (not shown) of pNC191. The 1.8 kbp *Sph*I/*Eco*RI fragment of pNC192 was then subcloned into pUC18 and pUC19 to generate the plasmids pNC184 and pNC194 respectively. Pigment production by pNC194 was at the same level as pNC191 and pNC192 but pNC184-containing transformants were Pig<sup>-</sup>. This observation suggested that the coding sequence of the enzyme or enzymes responsible for pigment production lay between the *Sph*I and *Eco*RI sites of the *Rhodococcus* DNA insert. This result also suggested that the pigment-producing insert of pNC181 was expressed from the putative *Rhodococcus* promoter which lay in the *Hind*III/*Sph*I fragment, the loss of which caused the Pig<sup>-</sup> phenotype of pNC184. The pigment-producing inserts of pNC191, pNC192 and pNC194, therefore, were expressed from the pUC *lacOP* promoter.

The plasmids pNC185 and pNC186 were made by subcloning the 2.1 kbp *Bgl*II fragment of pNC192 in both orientations into the compatible *Bam*HI site of pUC18. The insert of pNC185 was Pig<sup>+</sup> and was in the correct orientation to the *lacOP* promoter of pUC18 whereas the insert of pNC186 was in the reverse orientation and was Pig<sup>-</sup>. This result suggested that the promoter activity of pNC181 originated in the *Hind*III/*Bgl*II fragment which was not part of the original *Rhodococcus* DNA insert but came from the internal coding sequence of the *Eco*RI gene of the cloning vector pLR591 during the subcloning of the *Hind*III/*Eco*RV fragment of pNIL260 into pUC18 and pUC19.

The plasmid pKS1 was made by subcloning the *Hind*III/*Eco*RI fragment of pNC194 into the KS version of "Bluescript" in the correct orientation to the *lacOP* promoter. The plasmid pKS3 was made by deletion of the *Bgl*II/*Bam*HI fragment at the 3' end of the pKS1 insert and was prepared for the purpose of preparing sequencing templates by exonuclease III shortening.



**Fig. 3.1** Restriction maps of inserts in plasmids used in the analysis and sequencing of the pigment gene. The following plasmids were derived from pNIL260; pNIL270 (Pig<sup>-</sup>), deletion of *Sma*I/*Eco*RV fragment; pNC181 (Pig<sup>+</sup>) and pNC191 (Pig<sup>+</sup>), *Hind*III/*Eco*RV fragment of pNIL260 subcloned into pUC18 and pUC19 respectively; pNC192 (Pig<sup>+</sup>), deletion of 3' end of pNC191 up to *Eco*RI site; pNC184 (Pig<sup>-</sup>) and pNC194 (Pig<sup>+</sup>), *Sph*I/*Eco*RI fragment (1.5 kbp) of pNC191 subcloned into pUC18 and pUC19 respectively; pNC185 (Pig<sup>+</sup>) and pNC186 (Pig<sup>-</sup>), *Bgl*II fragment (2.1 kbp) of pNC192 subcloned into pUC18 in both orientations; pKS1 (Pig<sup>+</sup>), *Hind*III/*Eco*RI fragment (1.5 kbp) of pNC194 subcloned into "Bluescript"KS; pKS3 (Pig<sup>+</sup>), deletion of *Bgl*II/*Bam*HI fragment (1.3 kbp) of pKS1; pBS1 (Pig<sup>-</sup>), *Bgl*II/*Sph*I fragment (0.7 kbp) of pNC185 subcloned into pUC18; pHSLac1 (Lac<sup>+</sup>), *Hind*III/*Sma*I fragment of pNC191 (1.6 kbp) in pMC1871 *lac* fusion; pBSLac1 (Lac<sup>-</sup>), *Bgl*II/*Sph*I fragment of pNC191 (1.4 kbp) in pMC1871 *lac* fusion. Arrows denote direction of transcription from *lacOP* in pUC-derived plasmids. Flanking open boxes indicate vector sequences; thin lines indicate insert *Rhodococcus* DNA; solid boxes indicate the probable coding region and the contiguous open box represents the rest of the apparent ORF.

The size of the average bacterial gene is about 1 kbp and so the size of the DNA insert of the smallest Pig<sup>+</sup> plasmid, pKS3 (1.5 kbp insert), suggested that a single cloned *Rhodococcus* gene was probably responsible for pigment production in *E. coli*.

### 3.3.2 Nucleotide sequencing

A large open reading frame (ORF) of 1,653 bp was detected in the *Rhodococcus* DNA insert sequence which began with a methionine ATG codon at position 299 of the 2,109 bp sequence (Fig. 3.2). Five other in-frame methionine ATG codons were present at positions 305, 458, 581, 791 and 815 within the ORF. Since long stretches of ORF occur randomly in bacteria with a high G+C content, presumably due to the lack of A and T nucleotides which are a predominant feature of termination codons, additional evidence is necessary to assign the correct start codons in such organisms. The smallest Pig<sup>+</sup> subclones, pKS1 and pKS3, contained inserts of 1.8 kbp and 1.5 kbp respectively and neither contained the first four ATG codons at positions 299, 305, 458 and 581 within the ORF. However, both pKS1 and pKS3 produced pigment in *E. coli* cells suggesting that the translation initiation codon was either at position 791 or 815. The purine-rich sequence, GGAGGAAA-3-ATG preceding the ATG codon at position 791 was a possible ribosome binding site (Shine and Dalgarno, 1974). The ATG methionine codon at position 815 was not preceded by a purine-rich sequence. Therefore the ATG codon at position 791 seemed to be the most likely start point for translation.



AGATCTTCACAAAGGAAGGCACGCCATGACAACTCGTTCCGCGCCCGCGGCCCAACGGTCATGGTCGACGATCGCCGCGAGGTAGCCGCGCCCTCGCCGCGGTGAGCTGATCA  
 50 100  
 TCCTCAGCCACCACCGCCGCGTGGTGGCTCGTCCGGTGGCGCGCGCGACACCGAGCAGATGGCCTTCCTGATCCGCCACACCCCGGATTCTGACAGGTGGCGCTGCACGAACGGA  
 150 200  
 TCTGTGACCGTCTCGTCTCTGCGGAAGCGACACCGGTGCGTCCCACTCACACTCGGCCATGGCAATGTGACGTCGACGCGCGAGTGACACCGGGATCTCCGCGCGCGACCGGCG  
 250 300 350  
 GCGCACGGCCCGCATCTCGCCGATCCGACACCGACCCCGCGGATCTCACCAGGCGAGGACCTGGTTCGGTGGCGGTGACACCCACAGAGTTTCGGATCGCTTTGACCACCGCGT  
 400 450  
 CAGTCGCCCTCCCGCTCACTGACGCCACTGCAACCGGAAGTGGGAGCGGTGTTCGCCGACCTCGACGGCATCGCCGACCCGACAGTGTCGGTGATCGACACGATCGCGAAATCCTCG  
 500 550  
 CCGACCGGTACGGGTCAACCTCTGCGCCGCTCGGACCCCGAGGACAGCGTCCGCTGACACCCCGCGGTGGCGCACACGACGACACCCGAAACCCACAGCATCCGCGAGCCCATG  
 600 650 700  
 GCGGTCCCGCATGCCCTGAGCAGCGCCGATCGCACCCACAGCACCAAGAAACAGCCACCTACACCGGAGGAAACA ATG GAC ATC ACC CGC ACC GAA CTC ATG GAC CGG 823  
 M D I T R T E L M D R  
 GTA CAC GCC CTG GTA CCG GCC TTC GCC GAA CGC GCA CAG AAG ACC GAA GAA AAC CGT GCT CCC CTC GAC GAG ACG ATC ACC GAC CTC ATC 913  
 V H A L V P A F A E R A Q K T E E N R A P L D E T I T D L I  
 GAC TCG GGA ATC CTG GCC ACG CTG ACC CCG AAG GAA TAC GGC GGT CTC GAA CTC GGG CTC GAC GTC GCC GCC GAC ATC GTC CGC ACC ATC 1 003  
 D S G I L A T L T P K E Y G G L E L G L D V A A A I V R T I  
 AGC GCC GTC TGC CCC TCG ACC GGC TGG GTC ACC TCC TTT TAC ATC GGA GCG GCC TGG CGG GTG AAC ATC TTC ACC GAA CAG GCC CAA CGG 1 093  
 S A V C P S T G W V T S F Y I G A A W R V N I F T E Q A Q R  
 GAA GTG TTC GCC GAC AAG CCC TAC ACC CTC ACC GCC GGC ACC GCG GCC CCG CTG GGG CAG GTG CAG AAG GTT GAC GGC GGC TAC CGC ATT 1 183  
 E V F A D K P Y T L T A G T A A P L G Q V Q K V D G G Y R I  
 ACC GGG CAG ACG GCA TGG AAC TCC GGG TCG GTG CAC GCG GAG TGG TTC ACC TTC GCC GGT GTG GTG TTC GAG GAG GGC AGC GCC CCC ACC 1 273  
 T G Q T A W N S G S V H A E W F T F A G V V F E E G S A P T  
 CCG TTG TGG TTC CTG GTG CCC CGC GAG GAC GTC AAG GTC CTC GAC ACC TGG TAC ATC GCC GGC ATG TCC GGC ACC GGC AGC AAT GAC ATC 1 363  
 P L W F L V P R E D V K V L D T W Y I A G M S G T G S N D I  
 AGC GTC GAC GAC GTG TTC GTC CCG GAG TAC CGC ACC GGG CCG TTC GCC CTC GCC CTC GCC GGT ACC GCT CCC GGG CAG CTG ATC CAC CCC 1 453  
 S V D D V F V P E Y R T G P F A L A L A G T A P G Q L I H P  
 AAC CCC ATG TAC CAC CTG CCG TTC CTG CCG TTC GCG ATG GCC GAG GTC ACC CCG GTC GTC GTC GGT GCC CTG CGC GGT GCG GCG GAC GCG 1 543  
 N P M Y H L P F L P F A M A E V T P V V V G A L R G A A D A  
 TTC GTC CAG CGC ACC AAG GAC CGT CAA GGA ACC ATC AGT CAG GAG AAG GCC TCC GGC AAG CAG GCC GCC CAG ATG CGC CTC GGC CGG GCG 1 633  
 F V Q R T K D R Q G T I S Q E K A S G K Q A A Q M R L G R A  
 CTG GCC GCC GCG GAC GCC GCC GAA ACC TTG CTC GAC GCC TTC TTC GAG CGC CTC ACG GCG CAG CGG CCC GAG CAG TCA GAC CCC CGT GAT 1 723  
 L A A A D A A E T L L D A F F E R L T A Q R P E Q S D P R D  
 CGT GCC GAA ATG AAG CTC AAG GCA GCA TAT CTG GCC GAT CTG GCC CGC AAT GCG CTC AAC GAC ATG GTG CGC GGA ATC GGT GGG GAC GGG 1 813  
 R A E M K L K A A Y L A D L A R N A L N D M V R G I G G D G  
 TAC CGG AAC TCC GCA CCG ATC CAG CGG TAC TTC CGT GAC CTG TCC GTG CTG TCG GTG CAC GCC TTC CTC GAC ATC GAC ACC GCC TCG GAA 1 903  
 Y R N S A P I Q R Y F R D L S V L S V H A F L D I D T A S E  
 ACG ATC GGC CGG TTC ACC CTC GAC CTG CCC GTC TCC GAC CCG CTC CTC TGA CGAGCTGTGGGATCAGGAGGAGGCGAAGTGGTGGGAGTAACGACGAG 2 006  
 T I G R F T L D L P V S D P L L \*\*\*  
 CGAACGCCTCGACGATGCAGCCAGAGACCTGCTGTTCGCGGGAGCGCGAACAGTCAACAACCTTCGCATCGACCCCGGTACGCGATGGCGAACTCGCCAGATCT\* 2 109

**Fig. 3.2** Nucleotide sequence of the *Rhodococcus* pigment gene and deduced amino acid sequence of the ORF. The potential Shine-Dalgarno sequence is marked as S-D. The first six potential ATG start codons are boxed. The glutamic acid residues, E, are indicated. An inverted complementary repeat sequence is indicated by opposing arrows. The stop codons are indicated by asterisks.

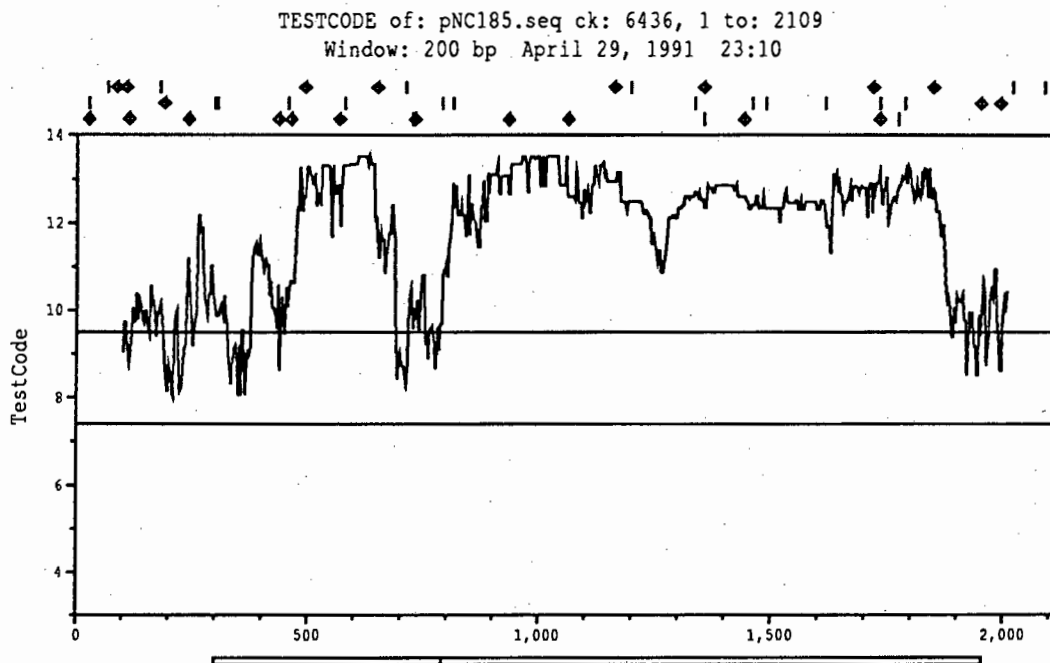
The 1,653 bp ORF was terminated by a TGA stop codon which was separated from a second stop codon, TAA, by 39 bp (Fig. 3.2). The proposed coding sequence within the 1,653 bp ORF, between the TGA stop codon and the ATG start codon at position 791, was 1,161 bp. The TGA stop codon is rarely used in *E. coli*, possibly because of its function of encoding the 21st amino acid, selenocysteine (Söll, 1988). An 8 bp stem, 4 base loop structure [ $\Delta G = -14.0 \text{ kcal mol}^{-1}$  or  $-58.6 \text{ kJ mol}^{-1}$ , (Salser *et al.*, 1977)] was present 78 bp downstream from the first stop codon. This stem-loop structure was followed by two T nucleotides although rho-independent termination sequences in *E. coli* normally contain at least four T nucleotides after the stem-loop structure (Rosenberg and Court, 1979; Brendel and Trifonov, 1984).

### 3.3.3 Codon usage and the ORF

Bacterial species which have a high G+C content, such as *Streptomyces*, show a strong bias towards the use of G or C in the third or "wobble" position of their codons (Bibb *et al.*, 1984). A low G+C content at the second position and an intermediate G+C level at the first position is also characteristic of the codon usage pattern of such bacteria (Bibb *et al.*, 1984). These features of the codon usage pattern have allowed the identification of *Streptomyces* protein-coding sequences (Bibb *et al.*, 1984).

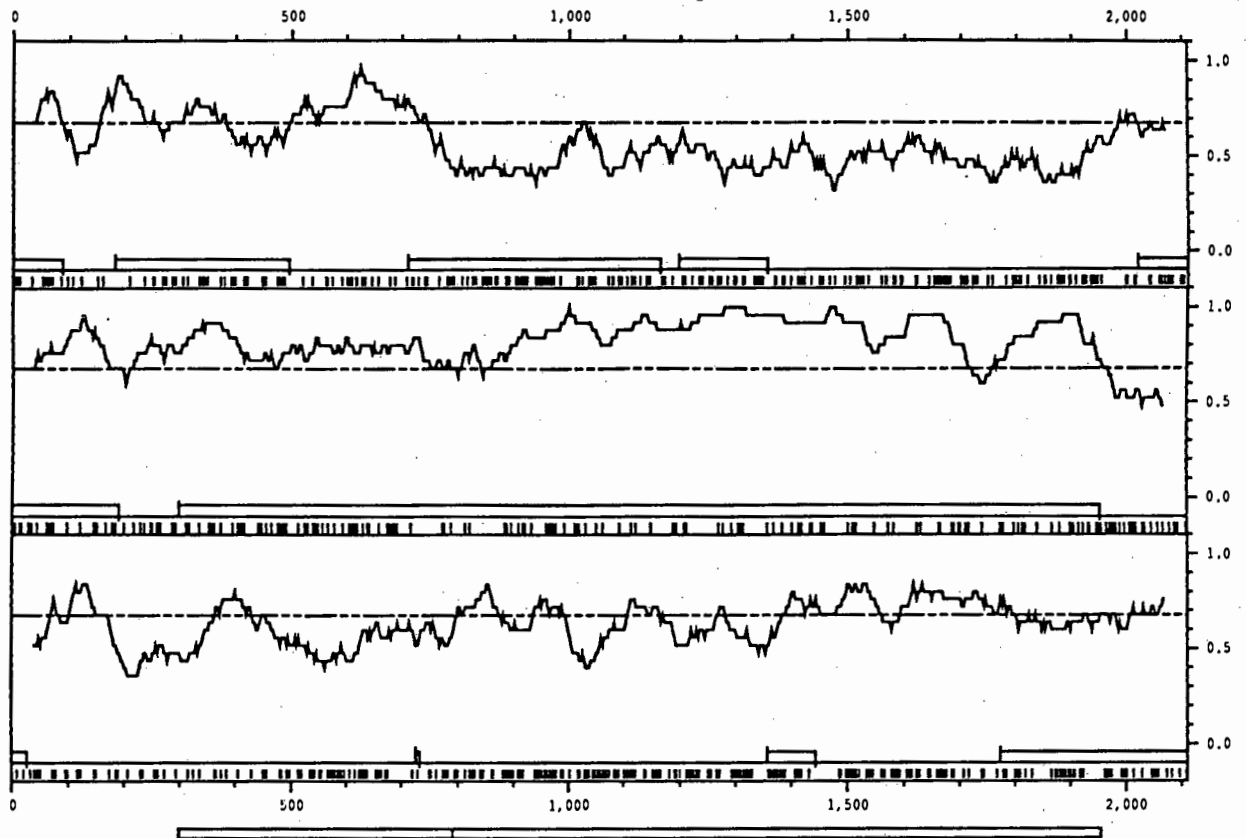
The 1,161 bp *Rhodococcus* DNA coding sequence which had an overall G+C content of 67.3 % was analysed for these codon usage patterns. A non-random distribution of G and C nucleotides started at the ATG methionine codon at position 791 bp and continued until the first stop codon, TGA at position 1,952 bp (Fig. 3.2). The overall G+C content of the 1,161 bp of the ORF starting at position 791 was 67.3 % and that of the first, second and third codon position was 66.9 %, 47.3 % and 88.1 % respectively. This nucleotide distribution is typical of bacteria with a high G+C content. The G+C distribution in the ten codons preceding the proposed start codon at position 791 was not greatly biased, with 60 %, 50 % and 40 % G+C for the first, second and third positions respectively, while the ten codons after the proposed

start codon had a definite G+C distribution bias of 60 %, 40 % and 90 % for positions one, two and three respectively. The strong bias towards G and C in the third position was only apparent after the start codon at position 791 and was not observed around the ATG methionine at position 815. The biased codon usage pattern was centred upon the start codon at position 791 bp, supporting the proposal that it was the translation start point. Analysis of the DNA sequence with the Testcode and Codonpreference programmes of the GCG computer analysis package, which both detect coding sequences from the G+C codon bias (Figs. 3.3a and 3.3b), graphically illustrate the position of the ORF.



**Fig. 3.3a** Analysis of the complete 2,109 bp nucleotide sequence of *Rhodococcus* DNA with the Testcode program of the GCG computer analysis package. Testcode detects protein coding sequences by plotting a measure of the non-randomness of the composition at every third base. The top window indicates probable coding sequences and the bottom window non-coding sequences, with a window of uncertainty in between. The 1,653 ORF and probable 1,161 bp coding sequence are indicated beneath the chart. The Testcode analysis indicates that a large ORF exists in the 1,161 bp sequence and also suggests that another, small coding sequence exists from about 400-650 bp.

CODONPREFERENCE of: pNC185.seq Ck: 6436, 1 to 2109 April 29, 1991 23:07  
 Codon Table: GenRunData:EcoHigh.Cod PrefWindow: 25 Rare Codon Threshold: 0.10  
 BiasWindow: 25 Density: 69.2



**Fig 3.3b** Analysis of the complete 2,109 bp nucleotide sequence of *Rhodococcus* DNA with the Codonpreference program of the GCG computer analysis package. Codonpreference can detect protein coding sequences by the bias of their G+C composition in the third position of each codon and it also detects the 1,161 bp coding sequence and the possible upstream ORF. However, Codonpreference also reveals sequences of rare codon usage (vertical lines) which suggests that the ORF between 400-650 bp is not a coding sequence due to the high prevalence of rare codons. The 1,161 bp ORF contains less densely packed rare codons, compatible with a coding sequence.

The pattern of codon usage illustrates the preference for codons ending in G or C in the wobble position (Table 3.2). This was true for all amino acids except glutamic acid where the codon GAA was used as frequently as GAG. The distribution of the two codons for glutamic acid was polarized; the GAA codon was predominantly utilized at the 5' end of the protein coding sequence and the GAG codon was preferred towards the 3' end. It is interesting that the occurrence of the GAA codon

was also greater than expected in the only other sequenced *Rhodococcus* gene, the nitrile hydratase gene (Ikehata *et al.*, 1989).

**Table 3.2** Codon usage in the protein coding region from position 791 bp to 1,954 bp.

|     |     |    |     |     |    |     |     |    |     |     |    |
|-----|-----|----|-----|-----|----|-----|-----|----|-----|-----|----|
| TTT | Phe | 1  | TCT | Ser | 0  | TAT | Tyr | 1  | TGT | Cys | 0  |
| TTC | Phe | 17 | TCC | Ser | 7  | TAC | Tyr | 9  | TGC | Cys | 1  |
| TTA | Leu | 0  | TCA | Ser | 1  | TAA | *** | 0  | TGA | *** | 1  |
| TTG | Leu | 2  | TCG | Ser | 5  | TAG | *** | 0  | TGG | Trp | 6  |
| CTT | Leu | 0  | CCT | Pro | 0  | CAT | His | 0  | CGT | Arg | 5  |
| CTC | Leu | 19 | CCC | Pro | 11 | CAC | His | 5  | CGC | Arg | 12 |
| CTA | Leu | 0  | CCA | Pro | 0  | CAA | Gln | 2  | CGA | Arg | 0  |
| CTG | Leu | 15 | CCG | Pro | 11 | CAG | Gln | 13 | CGG | Arg | 8  |
| ATT | Ile | 1  | ACT | Thr | 0  | AAT | Asn | 2  | AGT | Ser | 1  |
| ATC | Ile | 16 | ACC | Thr | 25 | AAC | Asn | 6  | AGC | Ser | 4  |
| ATA | Ile | 0  | ACA | Thr | 0  | AAA | Lys | 0  | AGA | Arg | 0  |
| ATG | Met | 8  | ACG | Thr | 5  | AAG | Lys | 10 | AGG | Arg | 0  |
| GTT | Val | 1  | GCT | Ala | 2  | GAT | Asp | 2  | GGT | Gly | 6  |
| GTC | Val | 14 | GCC | Ala | 33 | GAC | Asp | 26 | GGC | Gly | 12 |
| GTA | Val | 2  | GCA | Ala | 5  | GAA | Glu | 11 | GGA | Gly | 4  |
| GTG | Val | 11 | GCG | Ala | 11 | GAG | Glu | 10 | GGG | Gly | 8  |

### 3.3.4 Upstream DNA sequence analysis with *lac* fusions

The *lac*-fusion, promoter probe-vector pMC1871 (Casadaban *et al.*, 1983) contained a cloning cassette in front of the *lacZ* gene without the control region of the promoter, operator, translation initiation region and first eight non-essential *lacZ* codons. The Lac<sup>+</sup> phenotype will be expressed by *E. coli* transformed with this plasmid if it contains a DNA insert with an in-frame coding sequence, a promoter and a ribosome binding site. In addition, the host *E. coli* strain must be *lacY*<sup>+</sup> for lactose uptake.

The reading frame of the pigment gene was known from sequencing and a correctly positioned ribosome binding site was located. In-frame *lac* fusions were made with the *Hind*III/*Sma*I fragment of pNC192 and the *Bgl*II/*Sma*I fragment of pNC192 in pMC1871 to make the plasmids pHSlac1 and pBSlac1 respectively (Fig. 3.1). The *Bgl*II/*Sma*I fragment of pNC192 contained the the pigment gene ORF as far as the *Sma*I site, and 790 bp of non-coding upstream *Rhodococcus* DNA which contained the ribosome binding site and a proposed *E. coli*-functional *Rhodococcus* promoter. The

*Hind*III/*Sma*I fragment contained the *Bgl*II/*Sma*I fragment plus a further 250 bp of upstream vector sequence derived from pLR591 which had been carried over during initial subcloning of the *Rhodococcus* pigment-producing insert from pNIL260 into pUC19 and pUC18.

The fusions pHSlac1 and pBSlac1 both gave the blue colour reaction on LB Tc agar plates containing X-Gal. However, this reaction is extremely sensitive, detecting levels of  $\beta$ -galactosidase below those required for growth on lactose minimal medium (Casadaban *et al.*, 1983). The colour reaction from the cleavage of X-Gal is so sensitive that even *E. coli* containing inefficient *lac*-fusions can be detected. Such inefficient fusions include those without good promoters but with the two coding sequences in phase, and out of phase fusions with good promoters. *E. coli* cells containing these fusions were inoculated onto lactose MacConkey plates containing Tc on which only *E. coli* colonies capable of fermenting lactose turn red, indicating the Lac<sup>+</sup> phenotype (Casadaban *et al.*, 1983). *E. coli* cells containing pHSlac1 produced red colonies whereas those containing pBSlac1 were white, indicating that the *lac* fusion in pHSlac1 contained an efficient promoter and pBSlac1, despite its positive reaction on X-Gal, did not.

This result provided further evidence, in addition to the observation of the phenotypes of pNC185 and pNC186, that the apparent *Rhodococcus* promoter activity of pNC181 was due to a pseudo-promoter in the *Hind*III/*Bgl*II fragment of pLR591, possibly created by the juxtaposition of different fragments of DNA during the subcloning procedures.

### 3.3.5 Deduced amino acid sequence and *in vitro* transcription-translation

The deduced amino acid sequence of the 1,161 bp coding sequence from the start codon at position 791 consisted of 387 residues, with a calculated  $M_r$  of 42,056. The major labelled protein produced in the *in vitro* transcription-translation experiments by the insert DNA in plasmids pNC192 and pKS3 had an apparent  $M_r$  of approximately 42,000 (Fig. 3.4) which agreed well with the predicted  $M_r$  from the deduced amino acid sequence.



**Fig. 3.4** Autoradiogram of the cell-free coupled transcription-translation polyacrylamide gel. Lanes: 1) "Bluescript" KS; 2) pKS3; 3) pNC191; 4) pNC185; 5) pUC18; 6) no DNA. The bands at  $M_r$  31,000 represent  $\beta$ -lactamase; those at  $M_r$  42,000 represent the pigment gene product; those at  $M_r$  60,000 are an artifact of the cell-free transcription-translation system (Amersham).

The deduced amino acid sequence of the pigment gene product was compared with other known sequenced proteins in the databases using a limitation of 10 matches in a window of 30 residues. However, no significant amino acid sequence similarities were detected. A separate data base of aromatic dioxygenases was compiled as other evidence (Chapter 2) indicated that the protein product belonged to this family of enzymes, but no significant areas of homology were found with any members of this database.

### 3.4 DISCUSSION

The 2,109 bp *Bgl*III pigment-producing *Rhodococcus* DNA fragment as contained in the plasmid pNC186 was sequenced in both directions (Fig. 3.2). A number of common features of prokaryotic genes were sought within the sequence. A large open reading frame beginning with the methionine codon ATG or GTG is a feature of most *E. coli* genes although reading frames starting with TTG and ATT have also

been observed (Gold *et al.*, 1981; Kozak, 1983). The Shine-Dalgarno sequence (Shine and Dalgarno, 1974), the ribosome binding site, is usually located 4 to 15 nucleotides upstream from the start codon. This purine-rich sequence is complementary to the 3' end of the 16S ribosomal RNA and usually contains the sequence 5'-GGAGGT-3'. The sequences GAGG and GGAG which are contained in the Shine-Dalgarno sequence are rarely used in the coding region of prokaryotes, probably to avoid internal translation starts (Shpaer, 1986). In addition, evidence suggests that the first 24 nucleotides of the coding sequence contain conserved sequences complementary to the 5' terminus of the 16S ribosomal RNA which may be involved in ribosome binding in an analogous way to the Shine-Dalgarno sequence (Petersen *et al.*, 1988). Dreyfus (1988) also reported that the recognition of genuine *E. coli* translation start signals, as opposed to false starts, consisting of a Shine-Dalgarno-like sequence followed at the correct position by an ATG or GTG, occurring by chance in the coding sequence, depends on a 35 nucleotide sequence that brackets the start codon from the -20 to +15 positions.

The nucleotide sequence of the *Rhodococcus* pigment gene was located within a large ORF of 1,653 bp. Although this was a single ORF, the structural gene probably started with an ATG start codon situated 492 bp downstream of the first ATG of the large ORF. The nucleotide sequence of the pigment structural gene therefore consisted of 1,161 bp with the coding capacity for a protein of  $M_r$  42,056. The conclusion that the pigment gene consisted of 1,161 bp was based on the following evidence:

- (1) The close agreement between the calculated  $M_r$  of the pigment gene product (42,056) and the  $M_r$  determined by SDS-PAGE (42,000).
- (2) The asymmetrical distribution of G and C nucleotides observed in the codon usage pattern, which allowed the accurate and reliable prediction of the protein-coding sequence and the start of the pigment gene.
- (3) The ATG start codon of the 1,161 bp sequence was the only ATG of the larger 1,653 bp ORF which was preceded by a putative ribosome-binding site.
- (4) Deletion of the first four ATG codons of the longer 1,653 bp ORF did not prevent initiation of translation of the pigment gene mRNA and expression of the Pig<sup>+</sup> phenotype.



The expression of the pigment gene in both orientations in various vectors in *E. coli* initially suggested that the gene was regulated by an upstream sequence of *Rhodococcus* DNA in *E. coli*. However, it was shown that the *Rhodococcus* upstream DNA sequence did not contain any *E. coli* promoter consensus sequences and was unable to promote the expression of the pigment gene in *E. coli*. Pigment production by *E. coli* cells containing plasmids with the pigment gene inserted in reverse orientation to the *lacOP* promoter was due to a false promoter in the neighbouring vector DNA sequence of pLR591 which had been subcloned together with the *Rhodococcus* DNA insert.

The codon usage pattern for glutamic acid did not show the G+C bias in the wobble position which was observed in the codon usage pattern for all other residues. The Glu codons GAA and GAG were present at the same frequency. The bias against the expected greater use of the GAG codon may have been to avoid resemblances to the Shine-Dalgarno sequence. This hypothesis was supported by the observation that GAA codons were especially prevalent at the 5' end of the gene where translation is initiated. It has been reported that the sequences GAGG and GGAG, which are contained in the Shine-Dalgarno sequence, appear infrequently in the coding region of prokaryotic genes (Shpaer, 1986).

The nucleotide sequence of the pigment gene indicated that a single gene was responsible for the two pigments isolated from pigment-producing recombinant *E. coli*. The pigment gene product appeared to be novel since no amino acid homology was detected with other known protein sequences in the GenBank and EMBL data bases.

The small size of the pigment gene recommended itself for use in the development of a chromogenic vector for the detection of cloned inserts by insertional-inactivation and for other instances where reporter genes find application.

## CHAPTER FOUR

### CONSTRUCTION OF AN INSERTIONAL-INACTIVATION CLONING VECTOR

|  |    |
|--|----|
| 4.1 INTRODUCTION   | 66 |
| 4.2 METHODS AND MATERIALS  | 70 |
| 4.2.1 Bacterial strains, plasmids, phasmids and media                | 70 |
| 4.2.2 Preparation of plasmid and phasmid DNA                         | 72 |
| 4.2.3 Subcloning procedures  | 72 |
| 4.2.4 Oligonucleotide primers  | 72 |
| 4.2.5 DNA sequencing   | 73 |
| 4.2.6 Site-directed mutagenesis                                      | 73 |
| 4.3 RESULTS  | 76 |
| 4.3.1 Construction of subclones for mutagenesis                      | 76 |
| 4.3.2 Preparation of pMa3 ssDNA                                      | 76 |
| 4.3.3 Construction of gdDNA  | 76 |
| 4.3.4 Transformation, segregation and screening for mutants          | 77 |
| 4.3.5 Construction of the insertional-inactivation cloning<br>vector | 78 |
| 4.3.6 Analysis of pSLH8  | 83 |
| 4.4 DISCUSSION   | 85 |

## CHAPTER FOUR

### CONSTRUCTION OF AN INSERTIONAL-INACTIVATION CLONING VECTOR

**SUMMARY:** The construction of the insertional-inactivation vector pSLH8, which expresses an indigo marker gene, is described. The steps in its construction involved the site-directed mutagenesis of the *Rhodococcus* pigment gene and its insertion into pUC18 in such a way that the pigment gene enzyme was translationally fused with the *lacZ'* encoded  $\alpha$ -fragment of pUC18. A multiple cloning cassette containing seven commonly used restriction sites was designed and inserted into the same reading frame between the pigment gene and *lacZ'* sequence. It was demonstrated that insertion of foreign DNA into the multiple cloning cassette inactivated the pigment gene preventing pigment formation in transformants. The indigo reporter gene system was expressed on LB agar plates and has no requirement for expensive substrates such as X-Gal required for the  $\beta$ -galactosidase system of the pUC series of vectors. In addition, the *lacZ*/X-Gal system requires specific mutant strains for insertional-inactivation whereas the indigo marker gene should work in all commonly used strains of *E. coli*.

#### 4.1 INTRODUCTION

This chapter describes the design, construction and use of a chromogenic, insertional-inactivation cloning vector using the *Rhodococcus* pigment gene as a marker.

Three main properties are desirable for plasmid cloning vectors:

(i) Small size (less than 10 kbp); (ii) a readily selectable phenotype; and (iii) single sites for a large number of restriction endonucleases, and those preferably in genes with a readily selectable phenotype such as antibiotic resistance.

The advantages of a small size include possible resistance to damage by shearing during plasmid isolation; the potential for a high copy number and a reduced chance of multiple restriction sites.

One of the most useful cloning vectors is pBR322 (Bolivar *et al.*, 1977a; 1977b). It is small (4,363 bp) with a copy number of about 15 molecules per transformed *E. coli* cell which can be amplified to 1,000-3,000 in the presence of a protein synthesis inhibitor such as chloramphenicol. It also carries two sets of antibiotic resistance genes either of which can be used as a selectable phenotype for cells containing the plasmid. Single cleavage sites for a number of restriction endonucleases are situated within the coding sequences for each marker. The resistance phenotype is destroyed by the insertion of a fragment of foreign DNA into these sites.

An inconvenient aspect of the use of pBR322 is the selection procedure for distinguishing recombinants. For example, if foreign DNA has been cloned into restriction sites within the  $Ap^r$  gene, transformants will initially be selected on agar plates containing tetracycline which are then replica-plated onto agar plates containing ampicillin. After incubation, some of the original colonies will regrow whereas others will not. Those that do, the great majority, will consist of cells that carry the parental pBR322 molecule with no DNA insert and those relatively rare ones that do not will contain recombinant plasmid molecules. A number of techniques exist for increasing the probability of obtaining recombinant molecules such as phosphatase treatment of the vector prior to ligation (Maniatis *et al.*, 1982) and, where possible, the use of two restriction sites for cloning with the small DNA fragment between the two restriction sites being removed by gel purification before ligation.

A more convenient and reliable method for the selection of recombinant plasmids involves the elimination of transformants which do not contain foreign DNA inserts by the use of positive selection vectors. For example, the vector pEcoR251 contains the *EcoR1* gene under the control of the  $\lambda P_R$  promoter, and the  $Ap^r$  gene (Botterman, 1986). The *EcoR1* gene is lethal unless inactivated by the insertion of foreign DNA into the *EcoR1* gene.

Another convenient method for identifying recombinant plasmids in transformants involves the use of vectors containing reporter genes which are inactivated by the insertion of foreign DNA. Reporter genes usually confer a visible phenotype to

transformants, such as pigmentation, which allows discrimination between recombinants and non-recombinants. An important example of this type of selectable marker is the *lacZ'* gene carried by the pUC series of vectors (Vieira and Messing, 1982). The *lacZ'* gene encodes the  $\alpha$ -fragment, the first 146 amino acids of the amino terminal end of  $\beta$ -galactosidase. Some strains of *E. coli* have a deletion of the *lacZ'*-encoded  $\alpha$ -fragment in the *lacZ* gene for  $\beta$ -galactosidase, the M15 deletion, which is complemented by the presence of the *lacZ'* gene of the pUC vectors so that functional  $\beta$ -galactosidase is produced. The *lacZ'* gene on the pUC series of vectors contains different, unique restriction endonuclease sites clustered together to form a multiple cloning cassette (mcc). The mcc is situated within the amino-terminal part of the  $\alpha$ -fragment encoding sequence of *lacZ'*, a region of the  $\alpha$ -fragment polypeptide sequence which tolerates amino acid substitutions. When foreign DNA is inserted into the pUC mcc, the *lacZ'* gene is inactivated and *E. coli* ( $\Delta$ M15) transformants containing recombinant plasmids have no  $\beta$ -galactosidase activity whereas plasmids with no DNA inserts confer the  $\text{LacZ}^+$  phenotype. The transformants are differentiated by inoculating onto agar medium containing IPTG (isopropylthio- $\beta$ -D-galactoside, a gratuitous inducer of  $\beta$ -galactosidase) and X-Gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside, a chromogenic substrate). Colonies of transformants with DNA inserts are  $\text{LacZ}^-$  and detected as white colonies whereas those without inserts are  $\text{LacZ}^+$  and cleave the X-Gal to form blue colonies.

Other *E. coli* insertional-inactivation cloning vectors based on other reporter genes have been developed, partly to overcome the requirement for *lac* mutant host strains and the high cost of X-Gal. These include pJOE810 (Altenbuchner, 1988) and pIF413 (Tseng *et al.*, 1990) which both contain the melanin operon and a multiple cloning site, and pABC1 which uses  $\alpha$ -amylase as a reporter gene (Ikuta *et al.*, 1990). Other reporter genes have been developed for use in gene regulation studies, including  $\beta$ -glucuronidase (Jefferson *et al.*, 1986) and the luciferase operon (Olsson *et al.*, 1988).

The *Rhodococcus* sp. ATCC 21145 gene for indigo production seemed to be a useful candidate for development into a reporter gene in an insertional-inactivation cloning vector because of its small size, the speed and intensity of the pigment production, the intracellular confinement of the pigment and the lack of substrate requirements, LB agar being sufficient for pigment production.

Site-directed mutagenesis was used to create a new *SphI* site at the presumed start codon at position 791-793 (Fig 3.2) of the *Rhodococcus* pigment gene which would allow the construction of translational fusions with the pigment gene. This required the substitution of three nucleotides to change the sequence from 5'-ACA ATG GAC-3' to 5'-AGC ATG CAC-3'. An insertional-inactivation cloning vector was designed which consisted of the mutated pigment gene subcloned into the multiple cloning cassette of pUC18 in frame with the *lacZ'* encoded  $\alpha$ -fragment. This would result in the pigment gene being expressed from the *lacOP* promoter and the *lacZ* ribosome binding site, producing a fusion protein consisting of the pigment-producing protein and the amino acid sequence encoded by the *lacZ'*  $\alpha$ -fragment 5'. Thus, insertion of foreign DNA into the multiple cloning site of this new vector should prevent translational expression of the pigment gene resulting in colonies which contain recombinant plasmids displaying the Pig<sup>-</sup> phenotype whereas those colonies which contain the parental vector with no foreign DNA would be Pig<sup>+</sup>.

## 4.2 METHODS AND MATERIALS

### 4.2.1 Bacterial strains, plasmids, phasmids and media

The strains of *E. coli* used in this study, listed in Table 4.1, were maintained on LB agar (Appendix B) although strains WK6 and WK6mutS were regularly inoculated onto MM agar (Appendix B) to select for the *pro* marker present on the F' episome. The antibiotics Ap (100 µg ml<sup>-1</sup>) or Cm (25 µg ml<sup>-1</sup>) were used to select for *E. coli* transformants.

Phasmids (plasmid/phage hybrids) pMa5-8 and pMc5-8 were used for site-directed mutagenesis experiments (Stanssens *et al.*, 1989). These vectors contain, in addition to a ColE1-type origin of replication, the f1 filamentous phage origin of replication. This makes it possible to package one specific strand of the phasmids into phage rods. Phasmid pMa5-8 confers resistance to Ap and contains an amber mutation in the *cat* gene. The complementary vector pMc5-8 confers resistance to Cm and contains an amber mutation in the *bla* gene.

A single round of mutation using this twin vector system (pMa/c) yields not only the desired mutation but simultaneously leads to the acquisition of Ap<sup>r</sup> or Cm<sup>r</sup> and the alternative resistance gene containing an amber mutation. This latter feature permits selection against the template strand in the following round of mutagenesis if required. Thus, these vectors allow sequential introduction of a series of mutations during multiple rounds of mutagenesis involving alternative selection for resistance to Ap or Cm. The amber mutation in the *cat* gene of pMa5-8 (Ap<sup>r</sup> Cm<sup>s</sup>) destroys the unique *Pvu*II site, and the amber mutation in the *bla* gene of pMc5-8 (Ap<sup>s</sup> Cm<sup>r</sup>) destroys one of the two *Sca*I sites. Therefore, a restriction digest may be carried out to test the configuration of a specific mutant. *E. coli* WK6 was used as a recipient for the phasmids and their derivatives. *E. coli* WK6mutS is a mismatch repair deficient strain that is unable to suppress amber mutations. ssDNA was obtained by infection of phasmid-containing *E. coli* WK6 with the helper phage M13K07 which results in routinely high yields of phasmid ssDNA (Stanssens *et al.*, 1989).

**Table 4.1** Bacterial strains and plasmids used in this study.

| STRAINS                    | GENOTYPE  | SOURCE                                 |
|----------------------------|---|--|
| <i>E. coli</i> 3.300       | <i>lacI22 λ<sup>-</sup> relA1</i><br><i>spoT1 thi-1</i>   | Pardee <i>et al.</i> , 1959            |
| <i>E. coli</i> JM105       | <i>supE endA sbcB15 hsdR4</i><br><i>rpsL thi Δ(lac-proAB)</i><br>F'[ <i>traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZ</i><br>ΔM15]       | Yanisch-Perron <i>et al.</i> ,<br>1985 |
| <i>E. coli</i> LK111       | <i>lacI lacZ ΔM15 lacY<sup>+</sup></i>  | Zabeau and Stanley,<br>1982            |
| <i>E. coli</i> WK6         | Δ( <i>lac-proAB</i> ) <i>galE</i><br><i>strA</i> /F' <i>lacI<sup>q</sup> lacZ ΔM15</i><br><i>proA<sup>+</sup>B<sup>+</sup></i>            | Stanssens <i>et al.</i> , 1989         |
| <i>E. coli</i> WK6<br>mutS | Δ( <i>lac-proAB</i> ) <i>galE</i><br><i>strA mutS::Tn10</i> /F' <i>lacI<sup>q</sup></i><br><i>lacZ ΔM15 proA<sup>+</sup>B<sup>+</sup></i> | Stanssens <i>et al.</i> , 1989         |
| PLASMIDS                   | PHENOTYPE   | SOURCE                                 |
| pKS3                       | Ap <sup>r</sup> Lac <sup>-</sup> Pig <sup>+</sup>   | This study                             |
| pKSmut                     | Ap <sup>r</sup> Lac <sup>-</sup> Pig <sup>+</sup>   | This study                             |
| pMa2                       | Ap <sup>r</sup> Cm <sup>s</sup>   | This study                             |
| pMa3                       | Ap <sup>r</sup> Cm <sup>s</sup>   | This study                             |
| pMa5-8                     | Ap <sup>r</sup> Cm <sup>s</sup>   | Stanssens <i>et al.</i> , 1989         |
| pMc5-8                     | Ap <sup>s</sup> Cm <sup>r</sup>   | Stanssens <i>et al.</i> , 1989         |
| pSLH1                      | Ap <sup>r</sup> Lac <sup>-</sup> Pig <sup>+</sup>   | This study                             |
| pSLH4                      | Ap <sup>r</sup> Lac <sup>-</sup> Pig <sup>+</sup>   | This study                             |
| pSLH8                      | Ap <sup>r</sup> Lac <sup>-</sup> Pig <sup>+</sup>   | This study                             |
| pSmu25                     | Ap <sup>s</sup> Cm <sup>r</sup>   | This study                             |

#### 4.2.2 Preparation of plasmid and phasmid DNA

Plasmid DNA was prepared from *E. coli* strains by the alkaline lysis method of Ish-Horowicz and Burke (1981) (Appendix A). Large scale preparations were purified by CsCl-EtdBr equilibrium gradient centrifugation (Maniatis *et al.*, 1982) (Appendix A). Preparations for DNA sequencing were additionally treated with RNase A



(Maniatis *et al.*, 1982) followed by a phenol/chloroform extraction (Maniatis *et al.*, 1982), and ethanol precipitation (Maniatis *et al.*, 1982).

#### 4.2.3 Subcloning procedures

Restriction endonucleases (Boehringer-Mannheim, Anglian Biotechnology or BRL) were used as described by Maniatis *et al.* (1982). Digested DNA fragments were separated in agarose gels (0.8 % w/v) containing EtdBr ( $0.5 \mu\text{g ml}^{-1}$ ) by electrophoresis ( $2.5 \text{ V cm}^{-1}$ ) in TBE buffer (Appendix B) (Maniatis *et al.*, 1982). Gel-purification of DNA fragments (Appendix A) was performed by excision of EtdBr-stained bands from low-melting point agarose (Seaplaque) gels (1 % w/v) which were run in TAE buffer (Appendix B) at  $2.5 \text{ V cm}^{-1}$ . DNA was then separated from the agarose with "GeneClean" (Bio 101, California, USA) according to the manufacturers instructions. DNA ligations (Maniatis *et al.*, 1982) were performed in a mixture containing 1 U of T4 DNA ligase, 2  $\mu\text{l}$  of 10 x ligation buffer (Appendix B) and water up to a volume of 20  $\mu\text{l}$ . The DNA products of the ligation reaction were transformed into competent cells (Appendix A). Transformants were selected on LB agar plates containing the appropriate antibiotics.

#### 4.2.4 Oligonucleotide primers

Primers were prepared by Professor D. P. Botes (Department of Biochemistry, University of Cape Town) with an Applied Biosystems 381A DNA synthesiser (Forster City, California, USA) using their reagents and solvents. Proposed primer sequences were analysed for significant stem-loop structures which could affect the annealing behaviour of the primer, using the DNATools computer program.

A 17-mer synthetic primer with the sequence 5'-CGAAGGCCGGTACCAGG-3', complementary to the 5' end of the pigment gene (position 832-849, Fig. 3.2), positioned 31 bp 3' to the pigment gene start codon, was designed for nucleotide sequencing into the region upstream of the pigment gene to verify constructs from subcloning and mutagenesis procedures.

A 34-mer oligonucleotide (5'- ACA CCG GAG GAA AGC ATG CAC ATC ACC CGC ACC G-3') containing three mismatches (bold type) was used to introduce three nucleotide substitutions in one round of site-directed mutagenesis using the phasmid vector system (Stanssens *et al.*, 1989).

A multiple cloning cassette was constructed with two complementary oligonucleotides mcc1 (5'-CTCTGCAGCTGGATCCGGATATCCA-3') and mcc2 (5'-AGCTTGGATATCCGGATCCAGCTGCAGAGAGCT-3'). When annealed the dsDNA molecule had ssDNA overhangs at either end which would hybridize to overhangs generated by *Hind*III and *Sac*I restriction endonucleases. The nucleotide sequence of the cloning cassette was also designed to maintain the same reading frame between the upstream *lacZ'* coding sequence and the downstream pigment gene when subcloned into the *Sac*I/*Hind*III sites of pSLH1 to create the plasmid pSLH8. The reading frame of the multiple cloning cassette sequence was selected to contain codons which occur frequently in highly expressed genes of *E. coli* (Wada *et al.*, 1990).

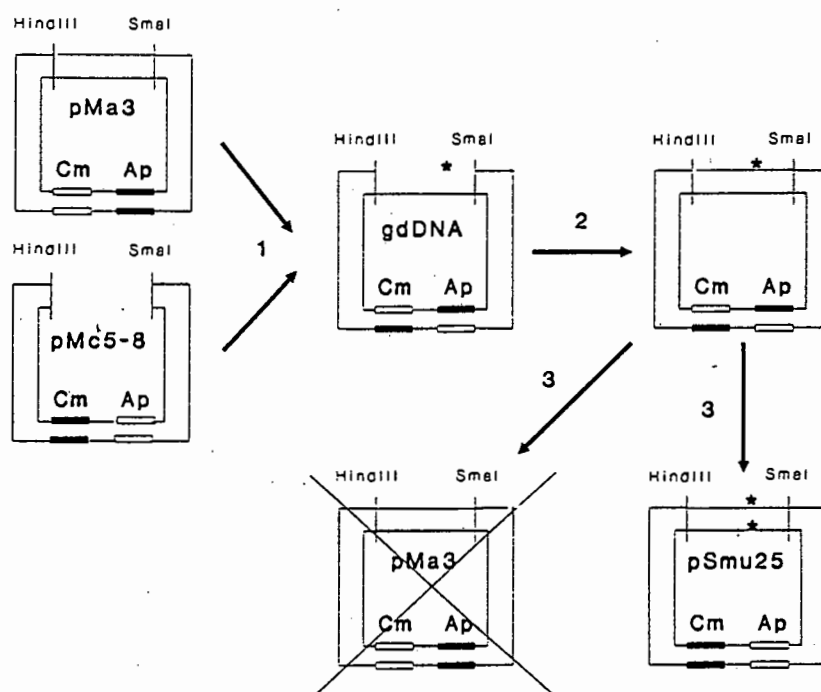
Annealing of the complementary oligonucleotides mcc1 and mcc2 was carried out in a mixture with approximately 250 pmol of each oligonucleotide, 2  $\mu$ l of 1.5 M KCl/0.1 M Tris-Cl, pH 7.5 and water up to a final volume of 20  $\mu$ l. This annealing mixture was heated to 60 °C for 5 min then allowed to cool slowly to room temperature. The dsDNA molecule generated was ligated (Maniatis *et al.*, 1982) into pSLH1 in an equimolar ratio of multiple cloning cassette : vector and transformed into competent cells of *E. coli*. Plasmid DNA was prepared from transformants and analysed for the presence of the multiple cloning cassette by restriction endonuclease analysis on agarose gels (Appendix A).

#### 4.2.5 DNA sequencing

DNA sequencing was done by the dideoxy chain-terminating method of Sanger *et al.*, (1977) with a "Sequenase" kit (US Biochemicals Corp.) following the manufacturers instructions. The DNA was radiolabelled with [<sup>35</sup>S]dATP [ > 1000 Ci (37 kBq) mmol<sup>-1</sup> ] (Amersham, UK).

#### 4.2.6 Site-directed mutagenesis

This procedure was performed essentially as described by Stanssens *et al.* (1989) using specially designed phasmids, pMa5-8 and pMc5-8, in a gapped duplex approach to site-directed mutagenesis (Kramer *et al.*, 1984) outlined in Fig. 4.1.



**Fig. 4.1** Site-directed mutagenesis procedure with the phasmid vector system (Stanssens *et al.*, 1989). Full explanation in text. The *Hind*III/*Sma*I fragment of pKS3 was inserted into the phasmid pMa5-8 ( $\text{Ap}^r$   $\text{Cm}^S$ ) and ssDNA prepared. The phasmid pMc5-8 ( $\text{Ap}^S$   $\text{Cm}^r$ ) was digested with *Hind*III and *Sma*I and denatured. **Step 1:** gdDNA was formed by hybridizing strands of DNA from pMa3 ssDNA and denatured pMc5-8. The mutagenic primer (asterisk) was annealed. **Step 2:** The gap was filled in and ligated and the strands of the resulting hybrid phasmid segregated by two rounds of transformation and phasmid DNA preparation. **Step 3:** Transformants containing the desired phasmid conformation, pSmu25 were selected for on Cm. Closed boxes signify functional genes and open boxes stand for defective genes.

(i) **Subcloning of target sequence.** Site-directed mutagenesis of gapped duplex (gdDNA) is more efficient if the gap is small and therefore a restriction fragment of only part of the pigment gene containing the target sequence was subcloned for mutagenesis. The fragment of DNA between the *Sma*I site located within the pigment gene in the plasmid pKS3 and the *Hind*III site of the multiple cloning cassette of pKS3 was subcloned into the *Sma*I/*Hind*III sites in the multiple cloning site of pMa5-8 to produce the phasmid pMa3 (Fig. 4.3) and transformed into *E. coli* WK6. Phasmid DNA was prepared from *E. coli* transformants and pMa3 was identified by restriction endonuclease digestion and agarose gel electrophoresis.

(ii) **Preparation of pMa3 ssDNA.** *E. coli* WK6 (pMa3) was inoculated into 5 ml LB Ap and incubated overnight at 37 °C and diluted 1:50 in 20 ml of fresh LB medium

without antibiotic. It was incubated at 37 °C until early log phase was reached, monitored by absorption at 600 nm, and infected with the helper phage M13KO7 at a m.o.i. of 20. After a 16 h incubation period, viral and pseudo-viral particles were recovered from the supernatant and the ssDNA was isolated by resuspending the phage pellet in 3 ml 50 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 (Kramer *et al.*, 1984) and extracted with phenol and chloroform (Maniatis *et al.*, 1982). The aqueous phase was adjusted to 0.3 mM sodium acetate and the ssDNA precipitated by addition of three volumes of ethanol. The yield of ssDNA was determined by UV-spectroscopy ( $\epsilon_{260} = 2.86 \times 10^{-2} \text{ cm}^2 \mu\text{g}^{-1}$ ) and was analysed on agarose gels.

(iii) **Preparation of the complementary phasmid.** The phasmid pMc5-8 was digested with the restriction endonucleases *HindIII* and *SmaI*, gel-purified and quantified by UV-spectroscopy. The linearised phasmid DNA was separated from the small restriction fragment by gel-purification followed by treatment with "GeneClean" as described previously. The yield of the large phasmid DNA fragment was estimated from EtdBr stained agarose gels.

(iv) **The construction of gdDNA.** A 36  $\mu\text{l}$  aqueous mixture of 0.1 pmol of the pMc5-8 fragment and 0.5 pmol pMa3 ssDNA was incubated at 70 °C for 5 min, 4  $\mu\text{l}$  of buffer containing 1.5 M KCl/100 mM Tris-HCl, pH 7.5, equilibrated to 70 °C was added and the mixture was allowed to cool to room temperature. The formation of gdDNA was monitored by agarose gel electrophoresis of a sample of the hybridisation mixture.

(v) **Annealing of the mutagenic primer.** The mutagenic 34-mer oligonucleotide was annealed to the gdDNA by the addition of 4-10 pmol (2  $\mu\text{l}$ ) of oligonucleotide to 8  $\mu\text{l}$  of the gdDNA mixture. This mixture was heated to 65 °C for 5 min and slowly cooled to room temperature. Four microlitres of 10 x fill-in/ligation buffer (625 mM KCl, 275 mM Tris-HCl, pH7.5, 150 mM  $\text{MgCl}_2$ , 20 mM DTT, 0.5 mM ATP and a 0.25 mM mixture of all four dNTP's), water to give a final volume of 40  $\mu\text{l}$ , 1 U DNA polymerase I (Klenow fragment) and 5 U of T4 DNA ligase were added and the mixture incubated at room temperature for 45 min.

(vi) **Transformation and segregation.** The polymerase/ligase reaction mixture was used to transform *E. coli* WK6mutS and a sample was spread on a selective medium to estimate the transformation efficiency. The remainder of the transformation mixture was used to inoculate 10 ml of LB medium containing Cm. After overnight

growth, phasmid DNA was isolated and the mixed phasmid population segregated by transforming *E. coli* WK6, again selecting for Cm resistance on LB agar plates (Fig. 4.1). Transformants were tested for sensitivity to Ap to avoid double transformants or parental pMac (Ap<sup>r</sup>, Cm<sup>r</sup>) configurations resulting from marker decoupling. Phasmid DNA was prepared from each candidate mutant and further biochemically tested with the restriction endonucleases *Pvu*II and *Sca*I. Finally, the presence of the mutation was confirmed by nucleotide sequencing using the 17-mer synthetic primer.

## 4.3 RESULTS

### 4.3.1 Construction of subclones for mutagenesis

The plasmid pMa3 was identified by *Eco*RI/*Sph*I restriction endonuclease digestion which generated a characteristic band of 750 bp on agarose gels. Plasmid pMa3 contained a restriction site for *Eco*RI in the phasmid multiple cloning cassette and a *Sph*I site in the insert whereas the parental pMa5-8 only contained the *Eco*RI site.

### 4.3.3 Preparation of pMa3 ssDNA

*E. coli* WK6 (pMa3) was infected with the helper phage M13KO7 causing the packaging of pMa3 ssDNA and M13KO7 ssDNA into phage rods. The total yield of ssDNA was estimated by UV spectroscopy to be 4.2 µg from 2 ml of bacterial culture.

### 4.3.3 Construction of gdDNA

The phasmid pMc5-8 was digested with the restriction endonucleases *Hind*III and *Sma*I and the yield of the linear vector minus the small fragment of the multiple cloning cassette was estimated to be 3 µg DNA in 20 µl TE. gdDNA was examined on agarose gels (Fig. 4.2). The mobility of gdDNA with small gaps is practically indistinguishable from that of relaxed fully double stranded plasmid DNA.



**Fig. 4.2** Mini-gel of phasmid DNA digestions of the steps in the site-directed mutagenesis procedure. Lane 1) lambda *Pst*I; 2) pMa3, undigested; 3) pMa3, *Sph*I/*Eco*RI; 4) pMa5-8, *Sph*I/*Eco*RI; 5) pMa3 ssDNA; 6) annealed pMa3 and pMc5-8 DNA to form gdDNA; 7) pMc5-8, *Hind*III/*Sma*I. Lanes 2-4 illustrate the formation of pMa3 and lanes 5-7 the formation of pMa3 ssDNA and gdDNA. Note the faint band, second from the top in lane 7 which probably represents the gdDNA.

#### 4.3.4 Transformation, segregation and screening for mutants

After the final transformation into *E. coli* WK6, forty transformants were screened for the presence of the mutation. Each colony was tested for the correct phasmid DNA configuration by reinoculating onto both LB Ap and LB Cm agar plates. All forty cultures regrew on Cm but 7 of these also grew on Ap and so were excluded from further analysis. Phasmid DNA was prepared from the remaining 33 transformants and the inserts identified by the analysis of the products of digestion with *Eco*RI and *Sph*I on agarose gels. The *Hind*III/*Sma*I restriction fragment initially subcloned from pKS3 into pMa5-8 already contained one *Sph*I site (Fig. 4.3). It was predicted that the introduction of a second *Sph*I site at the mutation target site would result in the reduction of the original 750 bp *Eco*RI/*Sph*I fragment of pKS3 and pMa3 to 690 bp in the mutation-containing phasmid derivative. Of the 33 DNA samples tested, 7 were found to produce the shortened *Eco*RI/*Sph*I fragment and therefore contained the mutated sequence. This was a mutation efficiency of 18 % which was lower than the average efficiency of 40-45 % reported by Stanssens *et al.* (1989), but the size and extent of the mutation performed in a single step in this experiment (3 bp in a 7 bp sequence) may have been responsible for the reduced

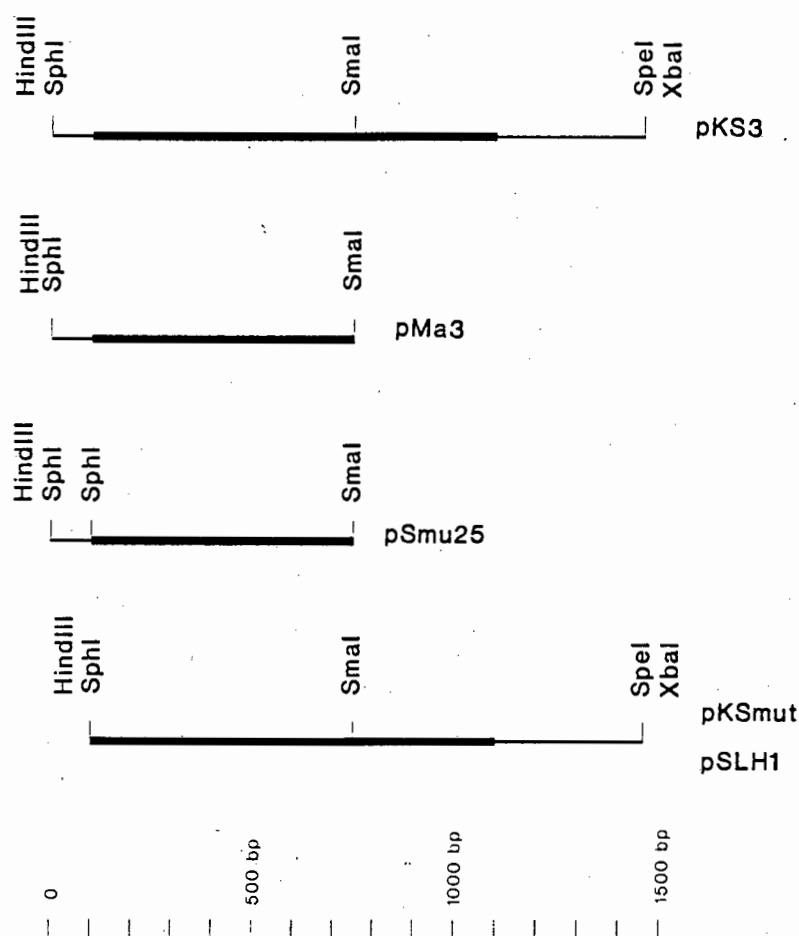
efficiency since the efficiency of annealing of the mutagenic primer to the ssDNA region in gdDNA is inversely proportional to the number of mismatched base pairs. One of the mutation-containing phasmid derivatives, designated pSmu25 (Fig. 4.3), was selected for further experimentation after confirmation of the construct configuration by analysis of the products of *ScaI* and *PvuII* restriction endonuclease digestion. The nucleotide sequence of the mutated region was determined by sequencing from the 17-mer primer. The new *SphI* site was detected at the predicted site.

#### 4.3.5 Construction of the insertional-inactivation cloning vector

The restriction fragment between the new *SphI* site and the internal *SmaI* site of the pigment gene of pSmu25 was subcloned into the *SmaI/SphI* sites of pKS3 from which the original *SmaI/SphI* fragment had been removed. The desired ligation product, designated pKSmut (Fig. 4.3), was identified after digestion at the *ClaI* and *XbaI* sites flanking the insert in the multiple cloning cassette and a comparison with the same digestion of pKS3 on agarose gels which yielded a larger restriction fragment.

It was predicted that the pigment gene would not be expressed in pKSmut because the pigment gene coding sequence lacked a ribosome binding site and was subcloned out of frame with the published nucleotide sequence of the *lacZ'* encoded  $\alpha$ -fragment of "Bluescript"KS. The lack of pigment production was intended to be a simple selection phenotype for pKSmut transformants compared with pigment-producing pKS3 transformants. However, *E. coli* LK111 transformed with pKSmut produced copious amounts of pigment on LB agar plates. The nucleotide sequence of the region immediately upstream of the pigment gene pKSmut was analysed by sequencing from the 17-mer primer. The presence of the mutated target sequence and the removal of the original ribosome binding site of the pigment gene was confirmed but two additional non-selected mutations were uncovered within the region upstream of the pigment gene in the "Bluescript" vector between the multiple cloning site and the start of the *lacZ'* encoded  $\alpha$ -fragment; a substitution of C to T at position 787 of the sequence published in the "Bluescript" instruction manual (Stratagene, San Diego, California) and a deletion of one of the three C nucleotides in positions 754-756. The deletion had the affect of restoring the pigment gene into the

correct reading frame, however, a TAA nonsense codon adjoining the *Hind*III site in pKSmut (nucleotide positions 718-720 of the "Bluescript"KS vector sequence) was also present in the new reading frame but did not prevent pigment production in *E. coli* LK111. The "Bluescript" vector mutations were not connected with the site-directed mutagenesis procedure but were probably acquired during routine strain and plasmid maintenance.



**Fig. 4.3** Phasmid and plasmid maps in the site-directed mutagenesis procedure showing relevant restriction sites. pMa3 was made by subcloning the *Hind*III/*Sma*I fragment of pKS3 (750 bp) into the phasmid pMa5-8. Note the formation of a new *Sph*I site at the start of the 1,161 bp coding sequence (black box) in pSmu25. pKSmut was made by subcloning the restriction fragment between the created *Sph*I site and the *Sma*I site of pSmu25 into the *Sph*I/*Sma*I sites of pKS3, effectively deleting the small *Sph*I fragment containing the original *Rhodococcus* ribosome binding site of the pigment gene. pSLH1 consisted of the pKSmut insert subcloned into pUC18 (see text).





after several days incubation at room temperature suggesting that LK111 is able to suppress TAA stop codons (Fig. 4.5).



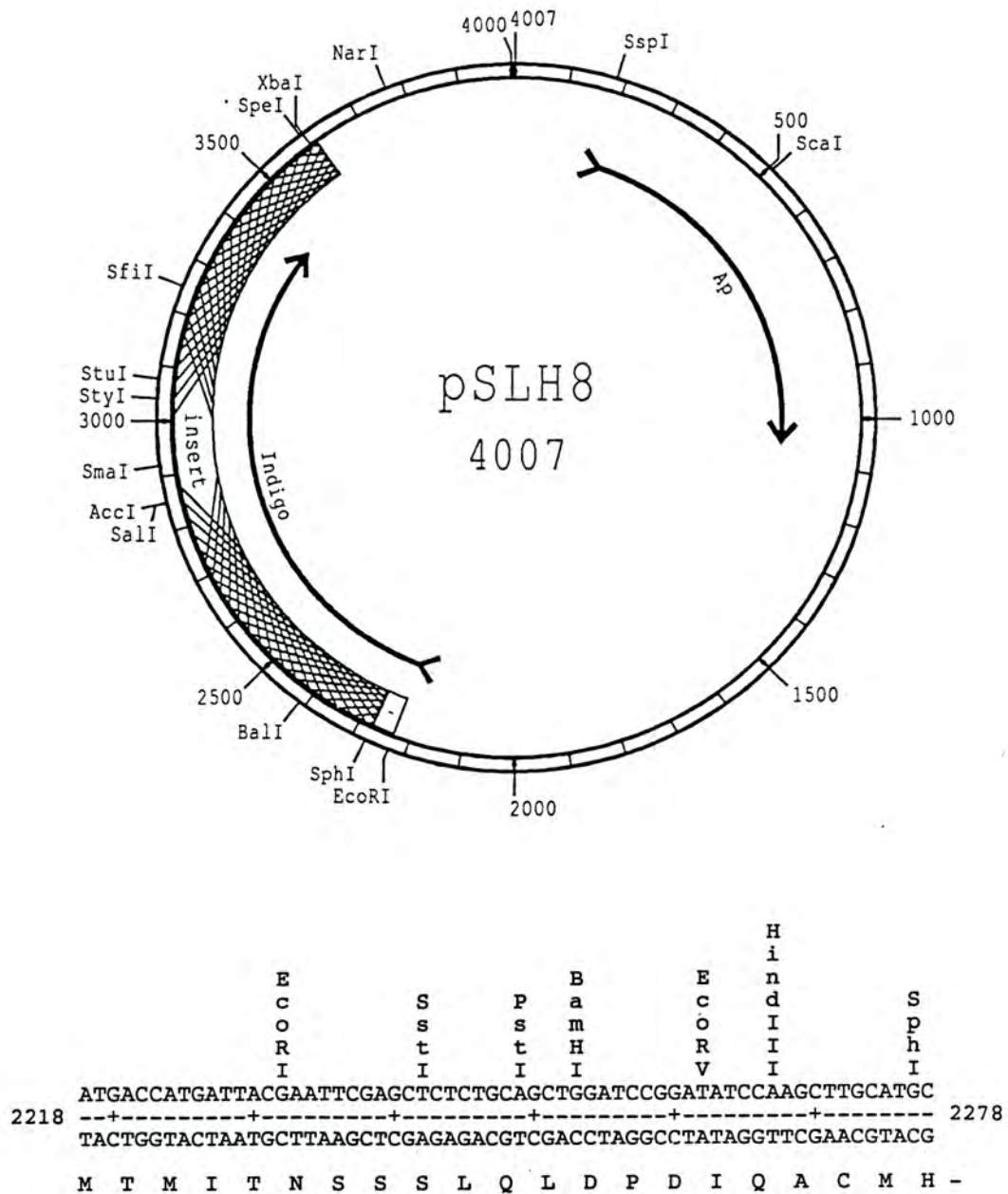
**Fig. 4.5** The plasmid pSLH1 contains a stop codon (TAA) in the pigment gene reading frame. *E. coli* LK111 (pSLH1) (right) is still capable of high levels of pigment production whereas *E. coli* 3.300 (pSLH1) (left) has a greatly reduced level of pigment production

The plasmid pSLH4 was generated by deletion of the *Pst*I/*Sph*I fragment in the multiple cloning site of pSLH1 to remove the in-frame termination codon and rare codons. An improved level of pigment production was observed by both *E. coli* strains LK111 and 3.300 transformed with pSLH4 compared to pSLH1. The fusion of part of the  $\alpha$ -fragment of  $\beta$ -galactosidase and the amino acid sequence encoded by the remaining part of the pUC multiple cloning cassette to the amino-terminal end of the *Rhodococcus* pigment-producing enzyme did not appear to affect the activity of that enzyme.

The sequence of the cloning cassette and further upstream regions of pSLH1 and pSLH4 was verified in each case by comparing the published sequence of the pUC18 *lacZ'*  $\alpha$ -fragment with the nucleotide sequence upstream of the pigment gene determined from the 17-mer primer.

The plasmid pSLH4, after the various genetic manipulations in its construction, contained a poor multiple cloning cassette of only three unique cloning sites, *Eco*RI, *Sac*I and *Bam*HI (Fig. 4.4). The plasmid pSLH8 (Fig. 4.6) was made by subcloning

the *mcc1/mcc2* synthetic multiple cloning cassette, which contained the restriction sites 5'-*EcoRI* *SacI* *PstI* *BamHI* *EcoRV* *HindIII* *SphI*-3' (Fig. 4.4), into pSLH1. Transformants containing the multiple cloning cassette were identified by the generation of single bands on agarose gels after digestion of plasmid DNA preparations with the restriction endonuclease *EcoRV*.



**Fig. 4.6** pSLH8 insertional-inactivation cloning vector showing *Rhodococcus* DNA insert (hatched box), pigment gene coding sequence (arrow) and  $Ap^r$  gene (arrow). The cloning cassette lies between the *SphI* and *EcoRI* sites.

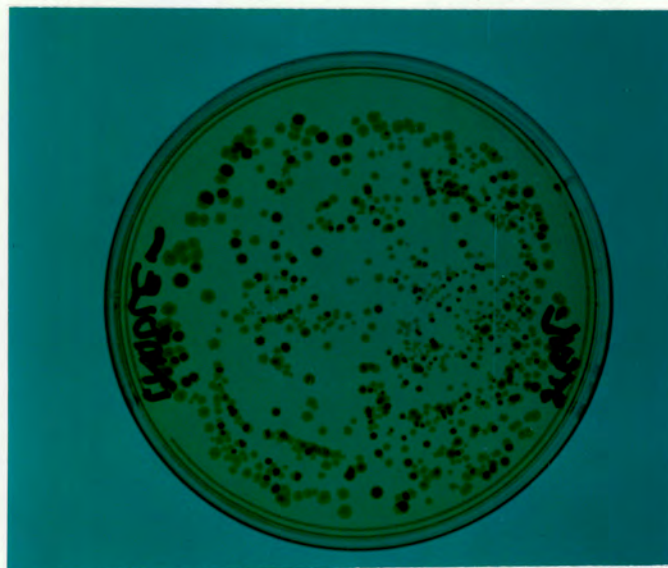


Four of twelve DNA preparations contained the single band and three of these four samples transformed *E. coli* to the Pig<sup>+</sup> phenotype. One of these DNA samples was sequenced with the 17-mer primer to confirm the correct placement of the multiple cloning cassette and a trial insertional-inactivation experiment was performed by randomly subcloning *Pst*I digested phage lambda DNA into the unique *Pst*I site in the multiple cloning cassette of pSLH8 (Fig. 4.4).

Transformants were a mixed population of blue and white colonies on LB Ap agar plates (Fig. 4.7). Plasmid DNA was prepared from six of each type of transformant and digested with *Pst*I restriction endonuclease. Agarose gel analysis of the digestion products showed that all of the blue pigmented transformants contained non-recombinant pSLH8 plasmid DNA whereas all of the plasmids from white transformants contained *lambda* DNA inserts of a wide range of sizes showing that pSLH8 was functioning as an insertional-inactivation cloning vector.

#### 4.3.6 Analysis of pSLH8

It was possible to reconstruct the complete nucleotide sequence of pSLH8 (Fig. 4.6) from the precise knowledge of the lineage of construction of this plasmid and the known nucleotide sequences of the pieces of DNA involved; the pigment gene, pUC18, "Bluescript"KS and the pSLH8 synthetic multiple cloning cassette.



**Fig. 4.7** LB Ap agar plate showing white (recombinant) and blue (non-recombinant) *E. coli* transformants, demonstrating insertional-inactivation of the pigment gene in pSLH8.

Apart from the restriction sites in the multiple cloning cassette of pSLH8, a number of useful, unique restriction sites were located within the pigment gene coding sequence for the detection of insertional-inactivation subcloning events:

Enzymes that cut pSLH8 in the multiple cloning cassette are:

|               |               |               |                 |
|---------------|---------------|---------------|-----------------|
| <i>Bam</i> HI | <i>Eco</i> RI | <i>Eco</i> RV | <i>Hind</i> III |
| <i>Pst</i> I  | <i>Sph</i> I  | <i>Sst</i> I  |                 |

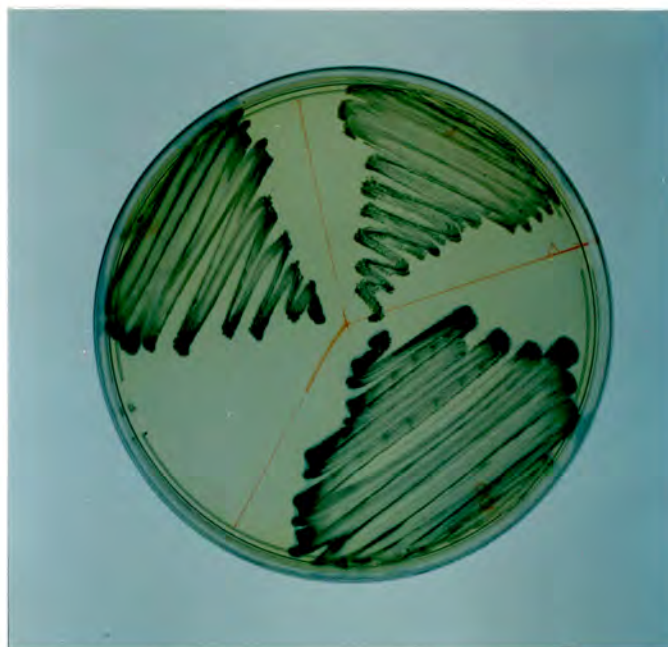
Enzymes that cut pSLH8 in the pigment gene once are:

|               |                |               |                  |
|---------------|----------------|---------------|------------------|
| <i>Acc</i> I  | <i>Bal</i> I   | <i>Bbv</i> I  | <i>Bsp</i> MII   |
| <i>Drd</i> II | <i>Eco</i> BI  | <i>Eco</i> KI | <i>Eco</i> R124I |
| <i>Mlu</i> I  | <i>Nru</i> I   | <i>Ppu</i> MI | <i>Sal</i> I     |
| <i>Sfi</i> I  | <i>Sgr</i> AI  | <i>Sma</i> I  | <i>Stu</i> I     |
| <i>Sty</i> I  | <i>Sty</i> SPI |               |                  |

The size of pSLH8 (Fig. 4.6) was determined from the compiled nucleotide sequence to be 4,007 bp. Large scale DNA preparations of pSLH8 from *E. coli* gave similar yields to pUC plasmids suggesting a high copy number of about 200 plasmid molecules per cell.

The rate of pigmentation of *E. coli* cells containing pSLH8 was estimated from colonies growing on LB Ap agar plates to be twice as fast as *E. coli* transformed with pSLH1 and indistinguishable from that of pSLH4 transformants (Fig. 4.8) suggesting that the pigment expression system of pSLH8 was working at an optimal level despite the fusion to the pigment-producing enzyme of a nineteen amino acid sequence encoded by the nucleotide sequence upstream of the pigment gene. Pigmented transformants were clearly distinguishable from non-pigmented colonies after overnight incubation at 37 °C followed by 8-20 h incubation at room temperature. The level of pigmentation deepened gradually over the next few days after further incubation at room temperature. The addition of 0.5-1 mM indole to the medium improved pigment production. Non-pigmenting colonies grew faster than pigmenting colonies, especially when 1 mM indole was included in the growth media suggesting that indigo rather than indole was the growth retarding factor.





**Fig. 4.8** Pigment production in *E. coli* LK111 by the insertional-inactivation cloning vectors. Top=pSLH1; bottom right=pSLH4; bottom left=pSLH8.

#### 4.4 DISCUSSION

A series of plasmid vectors pSLH1, pSLH4 and pSLH8 were constructed and tested for their ability to confer pigment production on *E. coli* and for the insertional inactivation of pigment production. All of these plasmids contained the *Rhodococcus* pigment gene, with a new *SphI* site created by site-directed mutagenesis at the ATG start codon, translationally fused to the *lacZ'*  $\alpha$ -fragment of pUC18. Plasmid pSLH1 also contained much of the cloning cassette of pUC18 which was useful for cloning and subcloning purposes but it also contained an in-frame nonsense codon and two very rare codons which greatly reduced the pigment-forming ability of cells transformed with this plasmid. These rare and nonsense codons were deleted from pSLH1 to form pSLH4 which had a much improved pigment producing capacity over pSLH1 but had lost most of the multiple cloning cassette. Therefore, a synthetic multiple cloning cassette was designed which featured seven unique, commonly used restriction sites plus a codon usage pattern characteristic of highly expressed genes in *E. coli*. This cloning cassette was ligated into pSLH1 to make the vector pSLH8 which was shown to confer high levels of pigment production on transformants. The insertion of foreign DNA into the multiple cloning cassette

resulted in pigment production being insertionally inactivated. Such transformants were white and clearly distinguishable from pigmented, non-recombinants within about 24 h of first inoculating transformation mixtures onto LB Ap agar plates.

The insertional-inactivation cloning vector pSLH8 fulfills the basic requirements for cloning vectors in that it is small, has several unique restriction sites and an easily selectable phenotype. It will be useful for many subcloning and cloning procedures. One of the advantages of this system is the ability of *E. coli* cells containing pSLH8 to produce the indigo marker on commonly used LB Ap agar plates without the requirement for any further media substrate supplementation.

All *E. coli* host strains transformed with pSLH8 should be capable of indigo production provided that they are wild-type for tryptophanase production which converts tryptophan to indole. This enzyme is expressed at a high level in most strains of *E. coli* although it is very sensitive to catabolite repression by glucose in the growth medium. Therefore, this indigo marker system will not be expressed on glucose-containing media. It is not yet known if the pigment gene will function as a marker in other bacteria.

The indigo marker system has two main advantages over the *lacZ'*  $\alpha$ -fragment complementation system of the pUC series of vectors. Firstly, the indigo marker system works on ordinary LB Ap agar plates with no additional substrate requirement whereas the *lacZ* system requires X-Gal, an expensive, artificial substrate which must be added to the surface of LB Ap agar plates or included in the media when pouring the plates. Secondly, the *lacZ* system requires special, *lacZ'*, mutant strains for the  $\alpha$  complementation system to work whereas the indigo system has a potentially much wider host range and should work in any *tna*<sup>+</sup> strain of *E. coli*.

An insertional inactivation vector using the melanin operon as a marker system (Tseng *et al*, 1990) is inferior to the indigo marker gene in that the melanin diffuses into the media and may make detection of recombinants difficult if large numbers of transformed colonies are close together on the surface of the agar whereas the indigo is almost entirely confined intracellularly. A modified medium supplemented with tyrosine and Cu<sup>2+</sup> ions was required for melanin production which was detectable after 2 days. In contrast, the indigo marker functions on LB or any other tryptone-containing medium and is detectable within 1 day.

This vector has properties suitable for use as an expression vector. The plasmid has a high copy number and inserts in the cloning cassette without their own promoters may be regulated by the upstream IPTG-inducible *lacOP* promoter, which also regulates pigment gene expression. These features result in high levels of gene expression and production of the protein product of the cloned gene.

This vector would also be useful for nucleotide sequencing. The high copy number ensures a ready source of the DNA sequence to be analysed and the vector sequence itself is already known, ensuring that vector sequence is not mistaken for the test sequence. The cloning cassette contains seven commonly used restriction sites three of which generate 3' overhangs and three 5' overhangs which would facilitate shortening of the DNA insert with exonuclease III (Henikoff, 1984). Sequencing primers are available which extend into the cloning cassette from either end so that inserts can be sequenced in both directions.

The fact that the pigment gene in pSLH8 is fused to an upstream coding sequence for nineteen amino acids and still functions suggests that the *Rhodococcus* pigment gene may also be useful in gene fusion studies in an analogous way to the *lacZ'*  $\alpha$ -fragment system (Casadaban, 1983). This would require the construction of a vector containing the pigment gene and the multiple cloning cassette without the upstream *lacOP* promoter or by subcloning the promoterless pigment gene downstream of other suspected promoter-containing pieces of DNA. The pigment gene can be removed from pSLH8 by digestion at any of the sites in the upstream multiple cloning cassette and either the *SpeI* or *XbaI* sites downstream of the pigment gene.

The small size of the pigment gene and the ability of its product to function in translational fusions suggests that the indigo marker gene has the potential for incorporation into a transposon such as Tn5 for the detection of insertional fusions in transposon mutagenesis experiments.

A promoter-probe vector incorporating the indigo reporter gene could be easily constructed. This would require a vector containing the pigment gene with a correctly positioned ribosome binding site, and the multiple cloning cassette with no upstream promoter and possibly a terminator upstream of the cloning cassette to prevent read through from distal promoter sequences.



## APPENDIX A

### STANDARD METHODS

|  |    |
|--|----|
| A.1 Large scale <i>E.coli</i> plasmid DNA isolation                                      | 89 |
| A.2 Plasmid purification by isopycnic CsCl-EtdBr density<br>gradient ultracentrifugation | 89 |
| A.3 Small scale isolation of <i>E.coli</i> plasmid DNA                                   | 90 |
| A.4 Restriction endonuclease digestions and DNA ligation reactions                       | 90 |
| A.5 Subcloning of gel purified restriction fragments                                     | 91 |
| A.6 Subcloning into vectors treated with calf intestinal<br>alkaline phosphatase         | 91 |
| A.7 Preparation of competent cells and their transformation                              | 92 |

## APPENDIX A

### STANDARD METHODS

#### A.1 Large scale *E.coli* plasmid DNA isolation

A 200 ml LB broth culture containing the appropriate antibiotic was grown overnight at 37 °C on an orbital shaker at 140 rpm. Cells were harvested by centrifugation at 1,600 × *g* for five min and resuspended in 4 ml of solution I (Appendix B). After standing at room temperature for 20 min, 8ml of solution II (Appendix B) was added and mixed by shaking then placed on ice for 5 min. Six ml of ice-cold solution III (Appendix B) was added and after mixing well stood on ice for about 10 min. The precipitate containing chromosomal DNA, proteins and SDS was removed by centrifugation (16,000 × *g* for 10 min). The supernatant fluid was mixed with an equal volume of propan-2-ol in a fresh centrifuge tube and left at room temperature for at least 2 min to precipitate plasmid DNA. A pellet was collected by centrifugation at 27,000 × *g* for 15 min which was washed in 70 % ethanol, resuspended in 5 ml of TE buffer and purified by isopycnic CsCl-EtdBr ultracentrifugation.

#### A.2 Plasmid purification by isopycnic CsCl-EtdBr density gradient ultracentrifugation

CsCl (final concentration, 1 g/ml) and EtdBr (final concentration, 125 µg/ml) was added to crude solutions of plasmid DNA in 5 ml of TE buffer. The refractive index of the solution was adjusted to 1.394-1.396 then centrifuged at 39,000 × *g* for 10 min to remove any undissolved debris. The supernatant fluid was sealed in a Beckman Quickseal ultracentrifuge tube and centrifuged at 340,000 × *g* for about 12 h in a Beckman vertical rotor. Plasmid bands were visualised by ultra-violet light (350 nm) and collected by piercing the side of the tube with a 1.2 mm guage needle attached to a syringe. The EtdBr was removed from the DNA sample by repeated extraction with an equal volume of salt-saturated propan-2-ol. Plasmid DNA was precipitated from the CsCl solution by adding two volumes of water followed by an

equal volume of propan-2-ol and standing at room temperature for at least two minutes. Pellets of plasmid DNA were collected by 15 min centrifugation in a microfuge then washed in 70 % ethanol and finally resuspended in TE buffer.

### **A.3 Small scale isolation of *E.coli* plasmid DNA**

A 5 ml LB broth culture of the *E.coli* strain containing the required plasmid was grown overnight at 37 °C in the presence of the appropriate antibiotic. Cells were harvested from 1.5 ml of the culture by centrifugation for 1 min in a microfuge. The cell pellet was drained of all LB broth, resuspended in 100 µl of solution I (Appendix B) and left at room temperature for 5-10 min followed by 1 min on ice. Solution II (Appendix B) was then added (200 µl) and mixed briefly on a vortex mixer. This mixture was returned to the ice bath for 5 min when 150 µl of solution III (Appendix B) was added. The mixture was again briefly vortex-mixed and left on ice for 5-10 min. The sample was then centrifuged in a microfuge for 5 min to remove all debris. The clear supernatant fluid was mixed with two volumes of 95 % ethanol in a fresh microfuge tube and the DNA precipitate was pelleted by centrifugation in a microfuge for 20 min. The pellet was dried and resuspended in 150 µl of TE buffer and then reprecipitated by the addition of 15 µl 3M sodium acetate, pH 5.0 and two volumes of 95 % ethanol. The samples were held at -20 °C for 20 min then centrifuged in the microfuge for 20 min. The pellets were again washed with 70 % ethanol, dried and resuspended in TE buffer (100 µl for high copy number plasmids such as pUC derivatives).

### **A.4 Restriction endonuclease digestions and DNA ligation reactions**

The procedures and precautions described by Maniatis *et al.* (1989) for the use of restriction endonucleases were followed. Digestions were usually carried out in a volume of 20 µl. In some cases the volume was increased so that the volume of restriction enzymes was never greater than 10 % of the final reaction volume (restriction enzymes are often delivered in a buffer containing glycerol which can have an inhibitory effect on the enzyme activity). Digestions were usually performed at 37 °C for 1 h in a buffer with a salt concentration which varied from 0-150 mM as specified by the manufacturers (restriction endonucleases were obtained from Boehringer Mannheim, Anglian Biotechnology and New England Biolabs). If

required, digestions were stopped by heating to 70 °C for 10 min or, when the restriction enzymes were stable to heat, by a phenol/chloroform extraction and propan-2-ol precipitation (Maniatis *et al.*, 1982).

Ligation reactions were usually performed at room temperature in a solution containing restriction endonuclease digested vector DNA and insert DNA, ligation buffer (Appendix B) and 1 U of T4 DNA ligase (Boehringer Mannheim) made up to 20 µl with sterile, distilled and deionised water.

#### **A.5 Subcloning of gel purified restriction fragments**

Restriction endonuclease digestion fragments were separated by agarose mini gel electrophoresis (1% Seaplaque low melting point agarose in TAE buffer). After staining the gel with EtBr the fragments were visualised by placing the gel on a 260 nm ultra violet transilluminator and the required fragment was excised with a fresh, sterile scalpel blade in as small a volume as possible. Agarose slices were melted at 70 °C for 5 min and the DNA fragment separated from the agarose with a "GeneClean" kit (Bio 101, La Jolla, California, USA) as described by the manufacturers. Briefly this procedure involves the adhesion of the DNA to a fine suspension of glass ("glassmilk") in the presence of a high concentration of NaI (which also helps to dissolve the agarose), pelleting the glassmilk and DNA, repeatedly washing the pellet with a 50 % solution of ethanol in a sodium chloride-EDTA buffer and finally eluting the DNA into a low salt buffer (such as TE) or water. The procedure is quick (20 min) and usually gives a high yield of the required fragment which is now ready for ligation.

#### **A.6 Subcloning into vectors treated with calf intestinal alkaline phosphatase**

Calf intestine alkaline phosphatase (CIP) removes 5'-phosphates from the ends of DNA restriction fragments so preventing their ligation. This treatment is useful to prevent the reformation of parental vector molecules in subcloning ligation mixtures.

Digested vector DNA was dissolved in CIP buffer (Appendix B) in a volume of 100 µl and CIP (Boehringer Mannheim, molecular biology grade) was added at a concentration of about 0.05 U for 1 pmole of restriction fragment ends (1 µg of a 3 kbp vector digested to give a single linear fragment has 1 pmole of ends). Digestion

with CIP was carried out at 37 °C for 30 min. Heat inactivation of CIP at 65 °C for 10 min stopped the reaction and the mixture was cleaned up with a single phenol/chloroform extraction, precipitated with ethanol, washed and resuspended in TE buffer before proceeding with the ligation reaction. The appropriate controls were included in the ligation experiment to assess the efficiency of the CIP reaction.

#### **A.7 Preparation of competent cells and their transformation**

This method is essentially that of Chung and Miller (1988). A fresh overnight LB culture of *E. coli* was used to inoculate a 20 ml flask of LB at a 1:50 dilution (400 µl of cell suspension in 20 ml LB) and grown to early log phase ( $OD_{600} = \pm 0.3$ ). Cells were decanted into SS34 centrifuge tubes (Sorvall), which had been pre-cooled on ice, and pelleted by centrifugation at  $1,600 \times g$  for 5 min at 4 °C. Cells were then resuspended in 2 ml of transformation and storage buffer (TSB) (Appendix B), pre-cooled on ice, and chilled on ice for 10 min. The plasmid DNA to be transformed (0.1-1000 ng) was placed in microfuge tubes on ice. Two control samples were always included; a no DNA sample to test for contaminants and a known amount of pBR322 (0.1 ng) to estimate the efficiency of transformation. 100 µl samples of the bacterial cells in TSB were mixed with the DNA samples and placed on ice for a further 10 min. 0.9 ml of TSB was then mixed with each sample and incubated at 37 °C for 1 h. 100 µl samples of each transformation mixture were spread on antibiotic-containing agar plates for selection of transformants. Transformation efficiencies of at least  $1 \times 10^7$  per µg of DNA were routinely obtained by this protocol.

## APPENDIX B

### BUFFERS AND MEDIA

|                                     |           |
|-------------------------------------|-----------|
| <b>B.1 SOLUTIONS AND BUFFERS</b>    | <b>94</b> |
| B.1.1 CIP Buffer                    | 94        |
| B.1.2 DNA loading buffer            | 94        |
| B.1.3 Ligation Buffer               | 94        |
| B.1.4 Plasmid preparation solutions | 94        |
| B.1.5 TAE Buffer                    | 95        |
| B.1.6 TBE Buffer                    | 95        |
| B.1.7 TE buffer                     | 95        |
| B.1.8 TSB                           | 95        |
| <b>B.2 MEDIA</b>                    | <b>95</b> |
| B.2.1 LB agar                       | 95        |
| B.2.2 LB broth                      | 96        |
| B.2.3 MacConkey agar                | 96        |
| B.2.4 MM agar                       | 96        |

## APPENDIX B

### BUFFERS AND MEDIA

#### B.1 SOLUTIONS AND BUFFERS

##### B.1.1 CIP Buffer (10 x)

|       |                   |
|-------|-------------------|
| 0.5 M | Tris-HCl, pH 9.0  |
| 10 mM | MgCl <sub>2</sub> |
| 1 mM  | ZnCl <sub>2</sub> |
| 10 mM | Spermidine        |

##### B.1.2 DNA loading buffer (5 x)

|              |                  |
|--------------|------------------|
| 0.25 % (w/v) | Bromophenol blue |
| 50 % (v/v)   | Glycerol         |
| 100 mM       | EDTA             |

##### B.1.3 Ligation Buffer (10 x)

|        |                   |
|--------|-------------------|
| 500 mM | Tris-HCl, pH 7.6  |
| 100 mM | MgCl <sub>2</sub> |
| 10 mM  | ATP               |
| 10 mM  | DTT               |

##### B.1.4 Plasmid preparation solutions (Ish-Horowicz and Burke, 1981)

|              |                       |
|--------------|-----------------------|
| Solution I   | 50 mM Glucose         |
|              | 25 mM Tris-HCl, pH8.0 |
|              | 10 mM EDTA, pH8.0     |
| Solution II  | 0.2 N NaOH            |
|              | 1 % SDS               |
| Solution III | 5 M Potassium acetate |

**B.1.5 TAE Buffer (50 x)**

|         |                     |
|---------|---------------------|
| 242 g   | Tris base           |
| 57.1 ml | Glacial acetic acid |
| 100 ml  | 0.5 M EDTA, pH 8.0  |

**Working solution:**

|         |              |
|---------|--------------|
| 0.04 M  | Tris-acetate |
| 0.002 M | EDTA         |

**B.1.6 TBE Buffer (5 x)**

|        |                    |
|--------|--------------------|
| 54 g   | Tris base          |
| 27.5 g | Boric acid         |
| 20 ml  | 0.5 M EDTA, pH 8.0 |

**Working solution:**

|         |             |
|---------|-------------|
| 0.089 M | Tris-borate |
| 0.089 M | Boric acid  |
| 0.002 M | EDTA        |

**B.1.7 TE buffer**

|       |                  |
|-------|------------------|
| 10 mM | Tris-HCl, pH 8.0 |
| 1 mM  | EDTA, pH 8.0     |

**B.1.8 TSB**

|       |                   |
|-------|-------------------|
| 10 %  | PEG 4000          |
| 5 %   | DMSO              |
| 10 mM | MgCl <sub>2</sub> |
| 10 mM | MgSO <sub>4</sub> |
|       | in LB broth       |

**B.2 MEDIA****B.2.1 LB agar**

LB broth with 1.5 % agar added.



**B.2.2 LB broth**

|      |                 |
|------|-----------------|
| 10 g | Tryptone        |
| 5 g  | Yeast extract   |
| 5 g  | NaCl            |
| 1 l  | Distilled water |

**B.2.4 MacConkey agar**

|      |                          |
|------|--------------------------|
| 40 g | MacConkey powder (Difco) |
| 1 l  | Distilled water          |

**B.2.4 MM agar****Salt solution (5 x):**

|        |                 |
|--------|-----------------|
| 52.5 g | $K_2HPO_4$      |
| 22.5 g | $KH_2PO_4$      |
| 5.0 g  | $(NH_4)_2SO_4$  |
| 2.5 g  | Sodium citrate  |
| 1 l    | Distilled water |

**MM agar:**

|       |                    |
|-------|--------------------|
| 15 g  | Agar (Oxoid No. 1) |
| 60 ml | Distilled water    |

Autoclave separately.

|        |                     |
|--------|---------------------|
| 200 ml | Salt solution (5 x) |
| 200 ml | Distilled water     |

Autoclave separately.

|       |  |
|-------|--|
| 10 ml | 20 % (w/v) carbohydrate                      |
| 1 ml  | $MgSO_4$                                     |
| 1 ml  | Vitamin B1                                   |
| 1 ml  | 20 % amino acid solution. Filter sterilised. |

## APPENDIX C

### Identity of blue and pink pigments isolated from *E. coli*: Preliminary study

|  |    |
|--|----|
| C.1 $^1\text{H}$ NMR spectrum of the blue pigment                    | 98 |
| C.2 $^1\text{H}$ and $^{13}\text{C}$ NMR spectra of the pink pigment | 99 |

## APPENDIX C

### Identity of blue and pink pigments isolated from *E. coli*: Preliminary study

by

Professor K. Koch, Department of Chemistry, University of Cape Town

#### C.1 $^1\text{H}$ NMR spectrum of the blue pigment

Figure 1a shows the full  $^1\text{H}$  NMR spectrum of the blue pigment. The low field region of the spectrum is shown expanded in Figure 2, from which it is clear that the resonance at 10.45 ppm may be associated with an N-H group, presumably due to an indole type structure. The remaining major resonances in the aromatic region consists of a triplet at 7.50 ppm and a doublet centred at 7.60 ppm and a resonance (resembling a doublet) at 7.69 ppm. The latter resonance integrates for *two* H atoms, relatively speaking. A homonuclear 2D correlation spectrum (COSY) shows positive correlations between the 6.94 ppm triplet and the 7.50 ppm triplet as well as the 7.60 ppm doublet (Figure 2). The 7.32 ppm doublet correlates with the 7.50 ppm triplet. Significantly, the resonance at 7.69 ppm does not show correlation peaks in this region, implying no significant coupling with other  $^1\text{H}$  atoms in the molecule. This latter resonance is reminiscent of a signal due to H(2) in the tryptophan molecule at 7.42 ppm, although the relatively large downfield shift of the former resonance at 7.69 ppm, means that an unambiguous assignment is not possible. Taken together, this  $^1\text{H}$  NMR data is consistent with an indole/tryptophan type structure although the hypothesis that the blue pigment is indigo cannot be confirmed with this data. If the resonance at 7.69 ppm is indeed part of the blue pigment molecule, then indigo must be ruled out in any event, since the two oxy-indole units are linked via the C2 atoms in indigo. The instability of the blue pigment and its tendency to convert into the pink form in DMSO-*ds* solution prevented the acquisition of a satisfactory  $^{13}\text{C}$  spectrum.

## C.2 $^1\text{H}$ and $^{13}\text{C}$ NMR spectra of the pink pigment

As seen from Figure 1b, the  $^1\text{H}$  NMR spectrum of the pink pigment is quite different from that of the blue pigment, particularly in the aromatic region. The  $^1\text{H}$  spectrum of the pink pigment consists of seven sets of resonances, four doublets, and three triplets. The triplet at 7.01 ppm has an integral twice that of the other two triplets, suggesting fortuitous resonance overlap (Fig. 3). Furthermore, two equally intense N-H type resonances are seen at 10.815 and 10.967 ppm suggesting two inequivalent indole-type spin systems, compatible with an indirubin structure. Unfortunately, the TLC pigment purification procedure left the pink component still mixed with some of the blue component which complicated interpretation of the  $^1\text{H}$  spectrum of the pink pigment, as seen in the difference spectrum of the two 1D  $^1\text{H}$  NMR spectra (Fig. 4).

Although only tentative evidence of a correlation between the aromatic and aliphatic region of the  $^1\text{H}$  spectrum has been found, the possibility of a chemical bond between a long chain alkyl fragment and the pink and blue chromophores cannot be ruled out. A tight host-guest type of molecular conjugate could also account for these spectra or inefficient purification of the pigments leaving other bacterial material in the pigment isolates.

The  $^{13}\text{C}$  NMR spectra of the pink pigment reveal the presence of at least 9 (possibly 10) resonances in the aromatic region (100-160 ppm) which also have an H atom attached. In addition, there are several resonances in the aliphatic region (0-50 ppm). The non-protonated quaternary and carbonyl carbon atoms which are likely to be present were not observed under the conditions tried (due to long relaxation times). To observe these signals, the spectra would have to be acquired for much longer using somewhat more concentrated solutions (Fig. 5). Given the data at hand and a HETCOR correlation spectrum, it is again possible to postulate two separate indole type ring systems (again, compatible with an indirubin structure) consisting of two sets of signals at 109.5, 121.3, 124.6 and 129.3 ppm as well as at 113.3, 121.4 (?), 124.3 and 128.5 ppm, corresponding to two sets of C-H carbon atoms in the benzene ring of an indole system. For comparison, the corresponding signals for indole are C5 (120.5 ppm), C6 (121.7 ppm), C7 (119.6 ppm) and C8 (111.0 ppm). The additional signal at 137.1 ppm in the spectrum of the pink pigment cannot be a quaternary carbon, since it is correlated with the  $^1\text{H}$  resonance at 8.74 ppm.

In conclusion, it is possible to state:

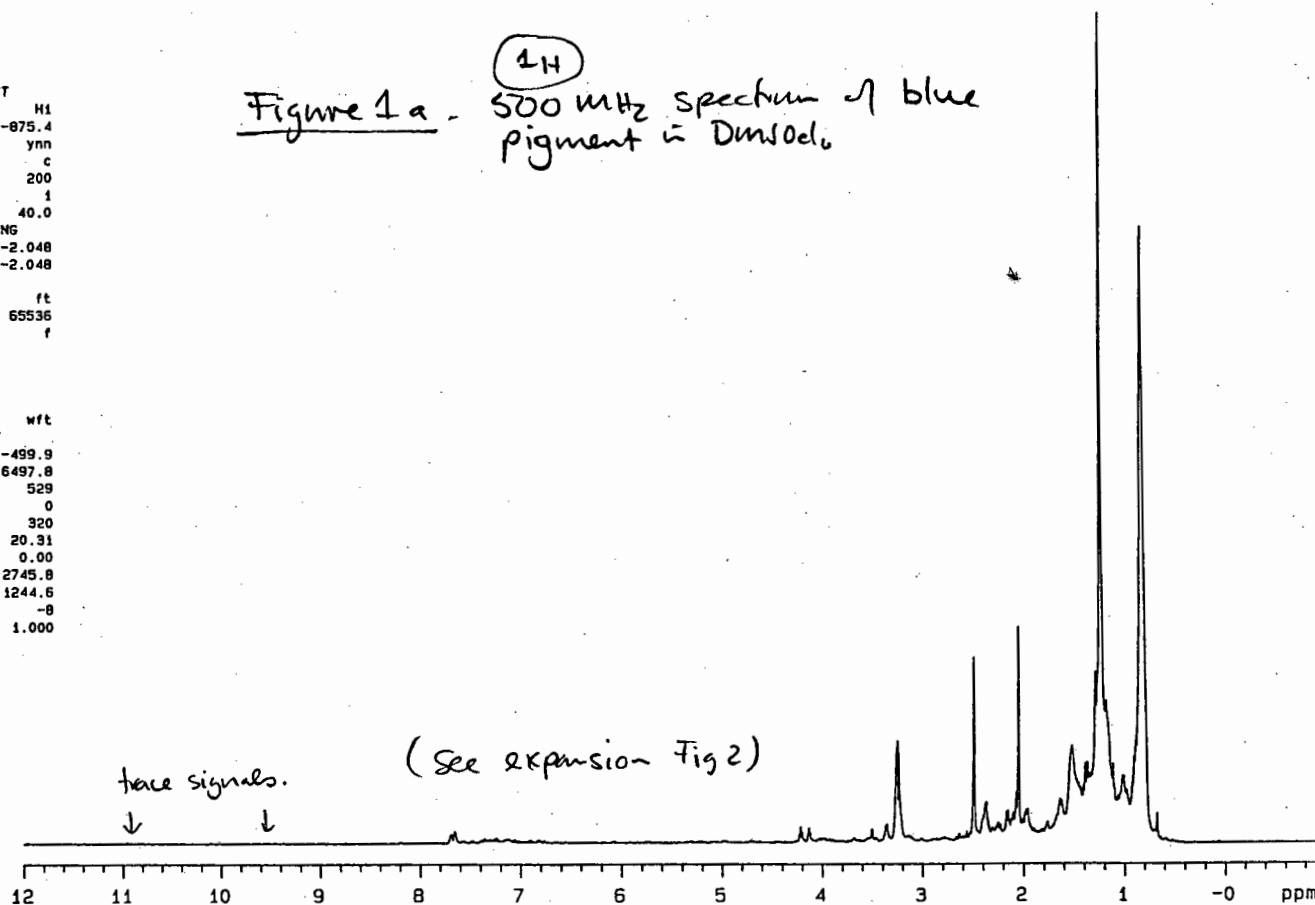
- i) The blue pigment  $^1\text{H}$  NMR spectrum has some features consistent with an indole type pigment, but it cannot be concluded that the pigment is indigo;
- ii) the blue pigment appears to change into the pink pigment, which is evident from the  $^1\text{H}$  spectrum of the latter, which also has the resonances of the blue pigment. Difference spectroscopy has confirmed this postulate;
- iii) the pink pigment has a fairly complex structure consisting of two indole-type structures linked in a manner which results in two separate spin systems being evident;
- iv) the role of the aliphatic residues in the pigments is unclear;
- v) the preliminary  $^1\text{H}/^{13}\text{C}$  NMR study of these pigments lend support to the hypothesis that these pigments are probably derived from tryptophan and are thus indole-type pigments. Nevertheless, it has not been conclusively demonstrated that the blue pigment is simply indigo.

Blue pigment

exp3 pulse sequence: s2pul

SAMPLE DEC. & VT  
 date Aug 21 90 dn H1  
 solvent DMSO dof -875.4  
 file exp dm ynn  
 ACQUISITION dmm c  
 sfrq 499.843 dmf 200  
 tn H1 dpwr 1  
 st 2.048 temp 40.0  
 np 32768  
 sw 8000.0 sb -2.048  
 fb 4400 sbs -2.048  
 bs 4 wtfile  
 tpwr 59 proc ft  
 pw 12.9 fn 65536  
 di 2.000 math f  
 tof 0  
 nt 128 verr  
 ct 128 wexp  
 alock n wbs  
 gain 0 wnt wft  
 FLAGS DISPLAY  
 il n sp -499.9  
 in n wp 6497.8  
 dp y vs 529  
 hs nn sc 0  
 wc 320  
 hzmm 20.31  
 is 0.00  
 rfl 2745.8  
 rfp 1244.6  
 th -8  
 ins 1.000  
 ei cdc ph

Figure 1a - <sup>1</sup>H 500 MHz spectrum of blue pigment in DMSO-d<sub>6</sub>



Nucleus H1 Freq. 499.84 MHz  
 Spec. Width 8000.00 Hz Offset 0 Hz  
 Acq. Time 2.05 sec Delay 2.00 sec  
 Pulse Width 12.90 sec Transmits 128.00

Nucleus H1 Offset -875.40 Hz  
 Mode ynn Power not used db  
 Modulation Mode c Freq. 200.00 Hz  
 Pulse Width 12.90 sec Power Mode not used

PLT/PROCESSING  
 FN 65536.00 RE sec CO sec  
 LB not used Hz AF sec CCO sec  
 Width 6497.76 Hz/ppm Start -499.94 Hz/ppm  
 Reference

EXPERIMENT  
 Pulse Sequence s2pul  
 Tube O.D. mm  
 Temp. 40.00 C  
 Solvent DMSO

SAMPLE  
 Blue pigment

Number  
 File exp  
 Date Aug 21 90  
 XL 500 V211

## STANDARD PROTON PARAMETERS

expt pulse sequence: s2pu1

SAMPLE DEC. & VT

date Aug 20 90 dn H1

solvent DMSO dof -873.2

file exp dm ynn

ACQUISITION dmm c

sfrq 499.843 dmf 200

tn H1 dpwr 1

st 0.256 temp 40.0

np 4096 PROCESSING

sw 8000.0 sb -0.256

fb 4400 sbs -0.256

bs 4 wfile

tpwr 59 proc ft

pw 12.9 fn 4096

d1 2.000 math f

tof 0

nt 4 werr

ct 4 wexp

elock n wbs

gain 0 wmt wft

FLAGS DISPLAY

il n sp -496.8

in n wp 6491.5

dp y vs 665

hs nn sc 0

wc 300

hzmm 21.64

is 501.07

rfl 2745.8

rfp 1244.6

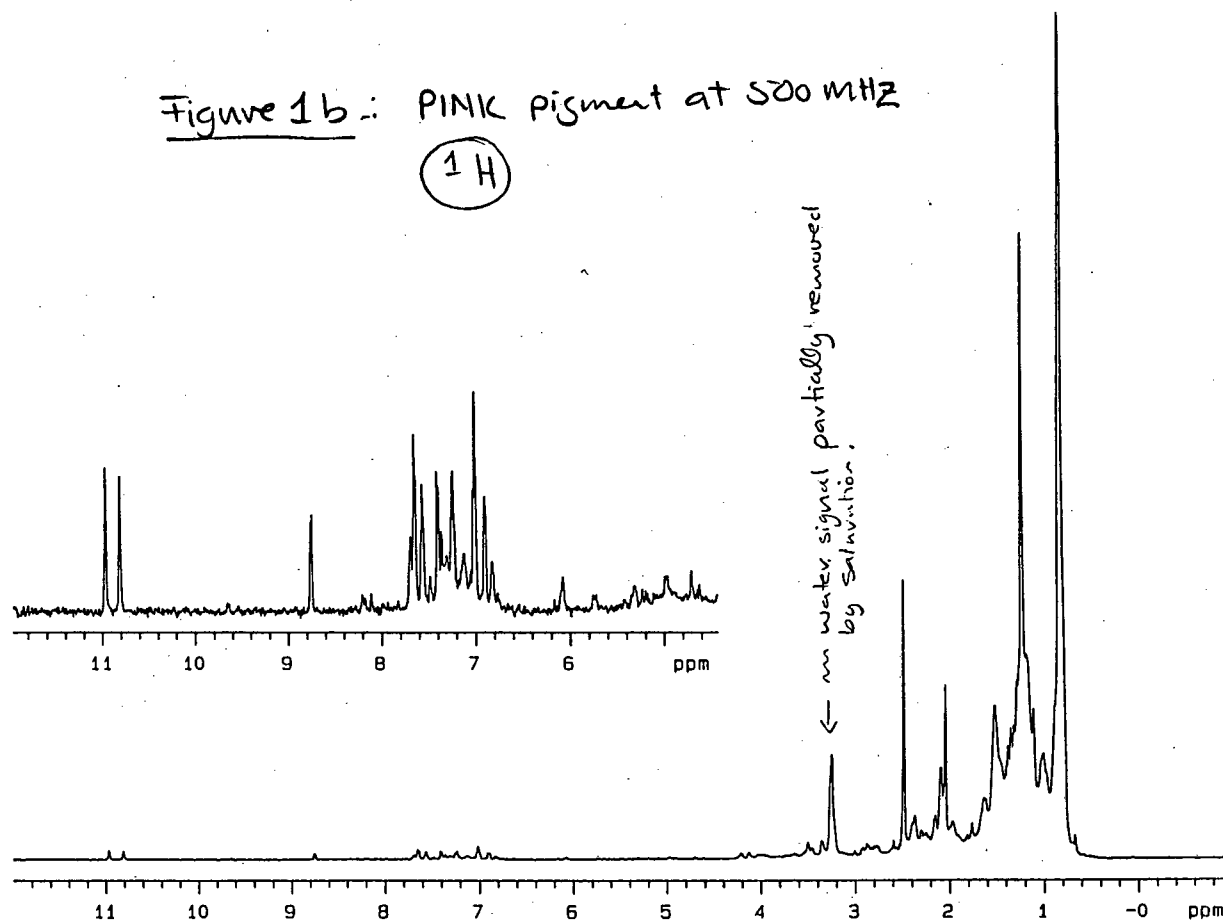
th -8

ins 1.000

ai cdc ph

Figure 1b: PINK pigment at 500 MHz

(1H)



Nucleus H1 Freq. 499.84 MHz

Soc. Width 8000.00 Hz Offset 0 Hz

Acq. Time 0.26 sec Delay 2.00 sec

Pulse Width 12.90  $\mu$ sec Transients 4.00

DECOUPLE

Nucleus H1 Offset -873.20 Hz

Mode ynn Power not used dB

Modulation: Mode c Freq. 200.00 Hz

Pulse Width         $\mu$ sec Power Mode not used

PLOT/PROCESSING

FM 4096.00  $\times$  RE        sec CO        sec

LB not used Hz AF        sec CCO       

Width 6491.45 Hz/ppm Start -496.79 Hz/ppm

Reference       

EXPERIMENT

Pulse Sequence s2pu1

Tube O.D.        mm

Temp. 40.00  $^{\circ}$ C

Solvent DMSO

SAMPLE STANDARD PROTON PARAMETERS

(pink pigment)

Number       

File exp

Date Aug 20 90


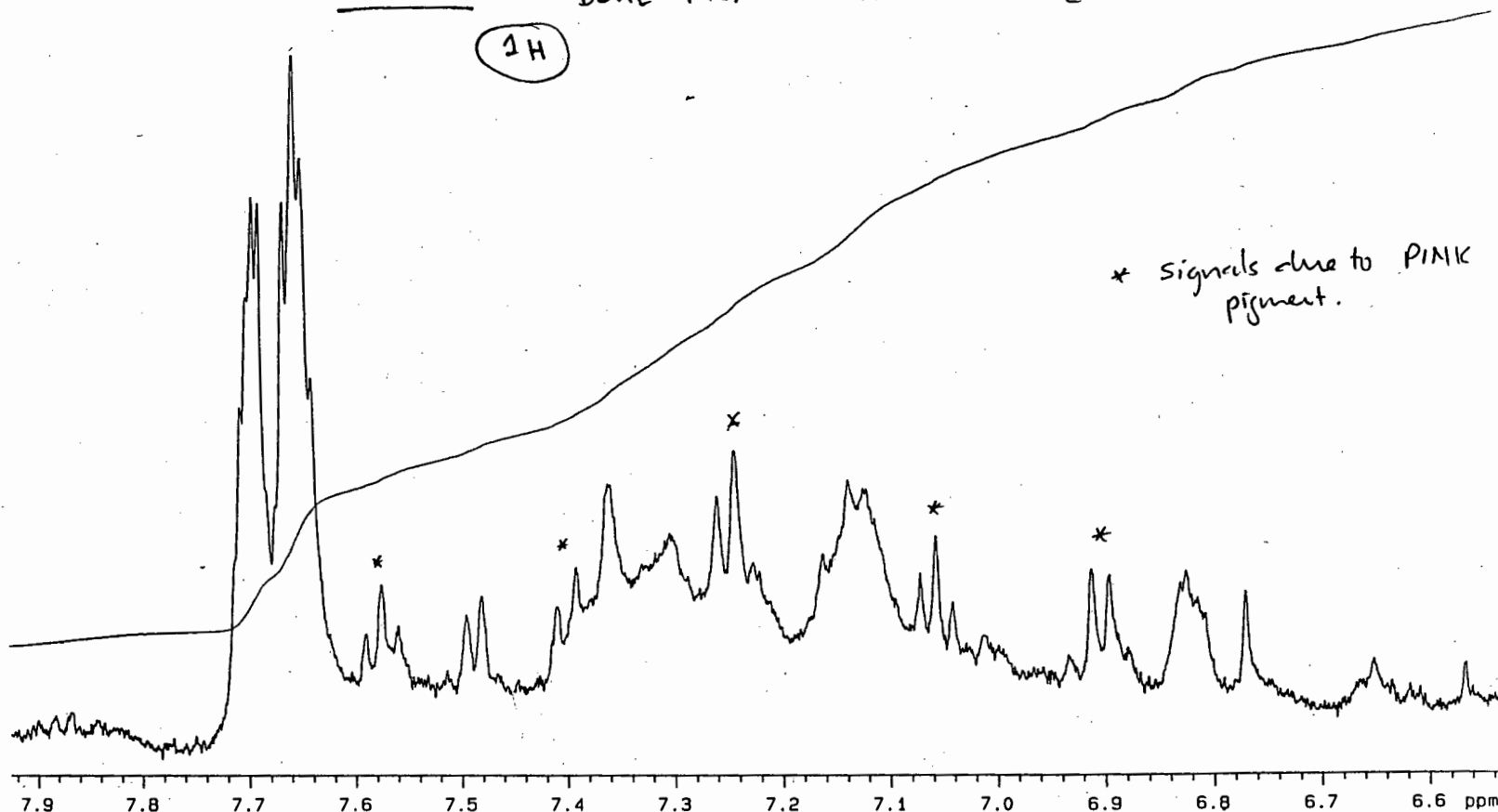
XL 500  varian

FIGURE 2

BLUE PIGMENT AT 500 MHz

(1H)



\* signals due to PINK pigment.

Nucleus H1 Freq. 499.84 MHz  
Spec. Width 8000.00 Hz Offset 0 Hz  
Acq. Time 2.05 sec Delay 2.00 sec  
Pulse Width 12.80  $\mu$ sec Transients 128.00

DECOUPLE Nucleus H1 Offset -875.40 Hz  
Mode ynn Power not used dB  
Modulation Mode C Freq. 200.00 Hz  
Pulse Width       $\mu$ sec Power Mode not used

PLT/PROCESSING F1 65536.00 K RE      sec CD      sec  
L1 not used Hz AF      sec CCO       
Width 704.61 Hz/ppm Start 3263.85 Hz/ppm  
Reference     

EXPERIMENT Pulse Sequence g2pu1  
Tube O.D.      mm  
Temp. 40.00 C  
Solvent DMSO

SAMPLE  
Blue pigment

Number       
File exp  
Date Aug 21 90  
XL 500 V211



FIGURE 2 BLUE PIGMENT 200 MHz (at uCT)

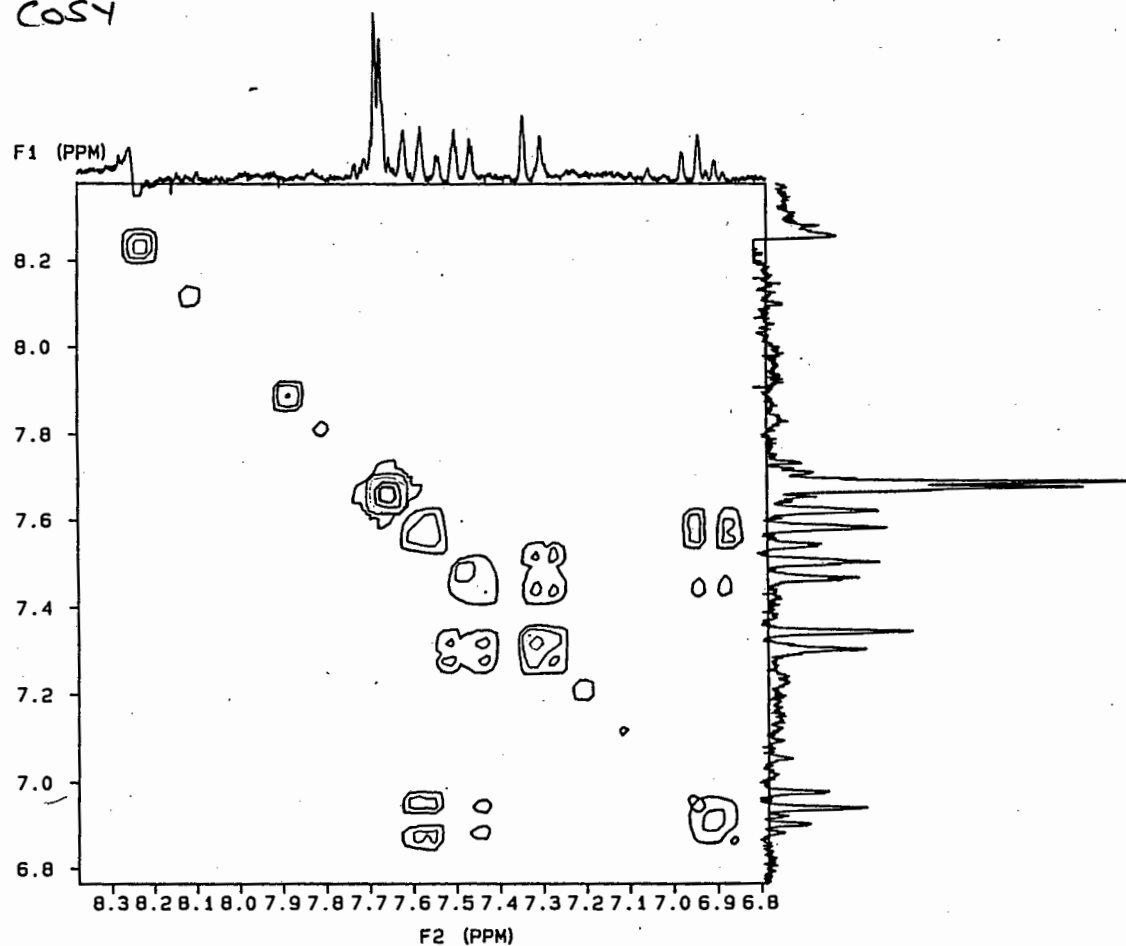
$$\textcircled{1_H} \quad \cos 4$$

MICRO S3 BLUE/DMSO T=40 31/6/90  
EXP4 PULSE SEQUENCE: COSY  
DATE 30-06-90  
SOLVENT DMSO  
FILE COSY

```

COSY PULSE SEQUENCE
OBSERVE PROTON
FREQUENCY 200.068 MHZ
1D SPECTRAL WIDTH (F2) 439.7 HZ
2D SPECTRAL WIDTH (F1) 439.7 HZ
ACQ. TIME 0.258 SEC
RELAXATION DELAY 1.0 SEC
PULSE WIDTH 90 DEGREES
FIRST PULSE 90 DEGREES
TEMPERATURE 40.0 C /
313.1 K
NO. REPETITIONS 128
NO. INCREMENTS 128
DATA PROCESSING
PSEUDO-ECHO SHAPED
FT SIZE 256 X 256
TOTAL TIME 6 HOURS
33.3 MINUTES

```

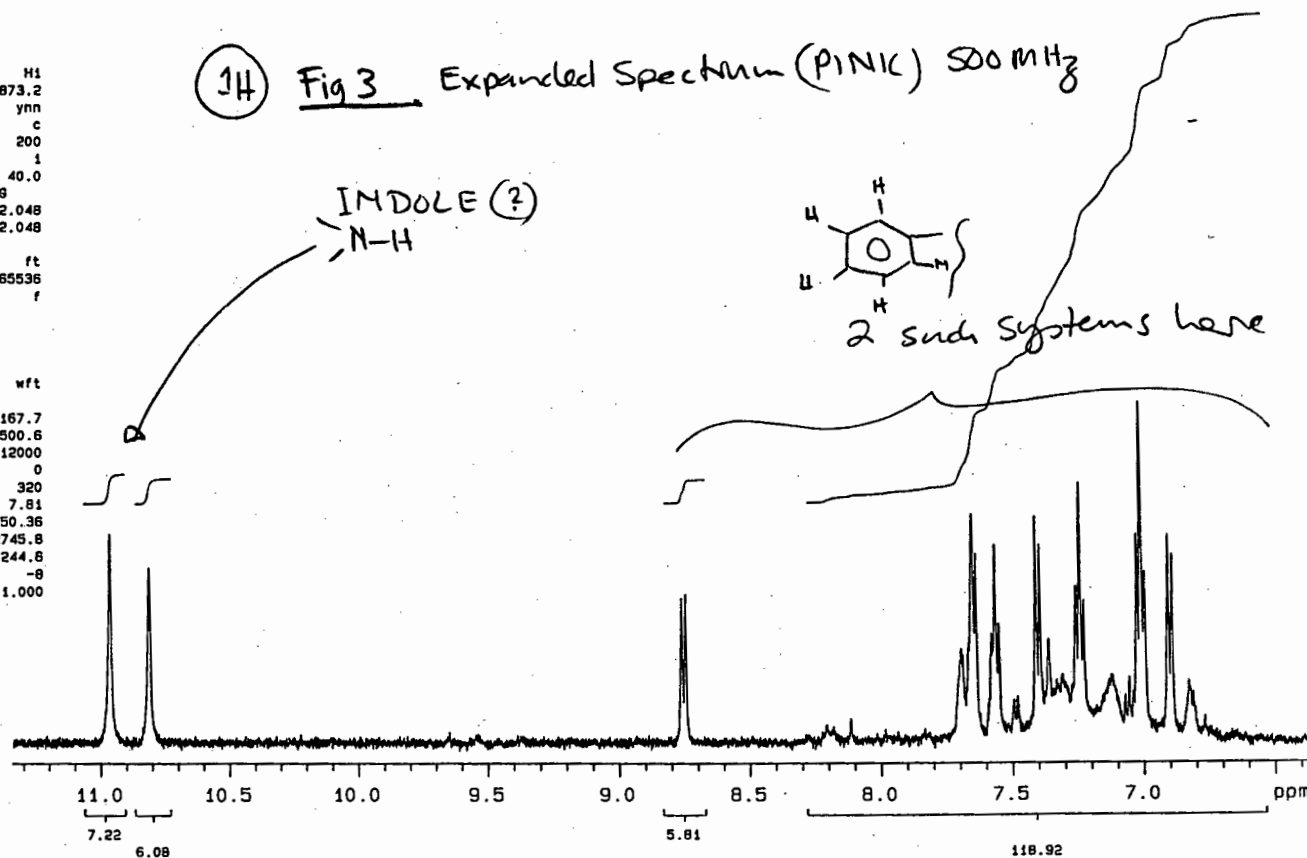


Pink pigment

exp3 pulse sequence: s2pul

SAMPLE DEC. & VT  
 date Aug 20 90 dn H1  
 solvent DMSO dof -873.2  
 file exp dm ynn  
 ACQUISITION dms c  
 sfrq 499.843 dmf 200  
 tn H1 dpwr 1  
 at 2.048 temp 40.0  
 np 32768 PROCESSING  
 sw 8000.0 sb -2.048  
 fb 4400 abs -2.048  
 bs 4 wtfile  
 tpwr 59 proc ft  
 pw 12.9 fn 65536  
 di 2.000 meth f  
 tof 0  
 nt 32 werr  
 ct 32 wexp  
 clock n wbs  
 gain 0 wnt wft  
 FLAGS DISPLAY  
 il n ap 3167.7  
 in n wp 2500.6  
 dp y vs 12000  
 hs nn ac 0  
 wc 320  
 hzm 7.81  
 la 25350.36  
 rfl 2745.8  
 rfp 1244.6  
 th -8  
 ins 1.000  
 ai cdc ph

14 Fig 3 Expanded Spectrum (PINK) 500 MHz



Nucleus H1 Freq 499.84 MHz  
 Spect. Width 8000.00 Hz Offset 0 Hz  
 Acq. Time 2.05 sec Delay 2.00 sec  
 Pulse Width 12.90 µsec Transients 32.00

DECOUPLE  
 Nucleus H1 Offset -873.20 Hz  
 Mode ynn Power not used dB  
 Modulation Mode c Freq 200.00 Hz  
 Pulse Width µsec Power Mode not used

PLT/PROCESSING  
 FN 65536.00 RE sec CD sec  
 LB not used Hz AF sec CCD  
 Width 2500.57 Hz/ppm Start 3167.8 Hz/ppm  
 Reference

EXPERIMENT  
 Pulse Sequence s2pul  
 Tube O.D. mm  
 Temp. 40.00°C  
 Solvent DMSO

SAMPLE  
 Pink pigment

Number  
 File exp  
 Date Aug 20 90  
 XL 500 V:111

Figure 4  $^{13}\text{C}$  NMR of PINK pigment (run at UCT)

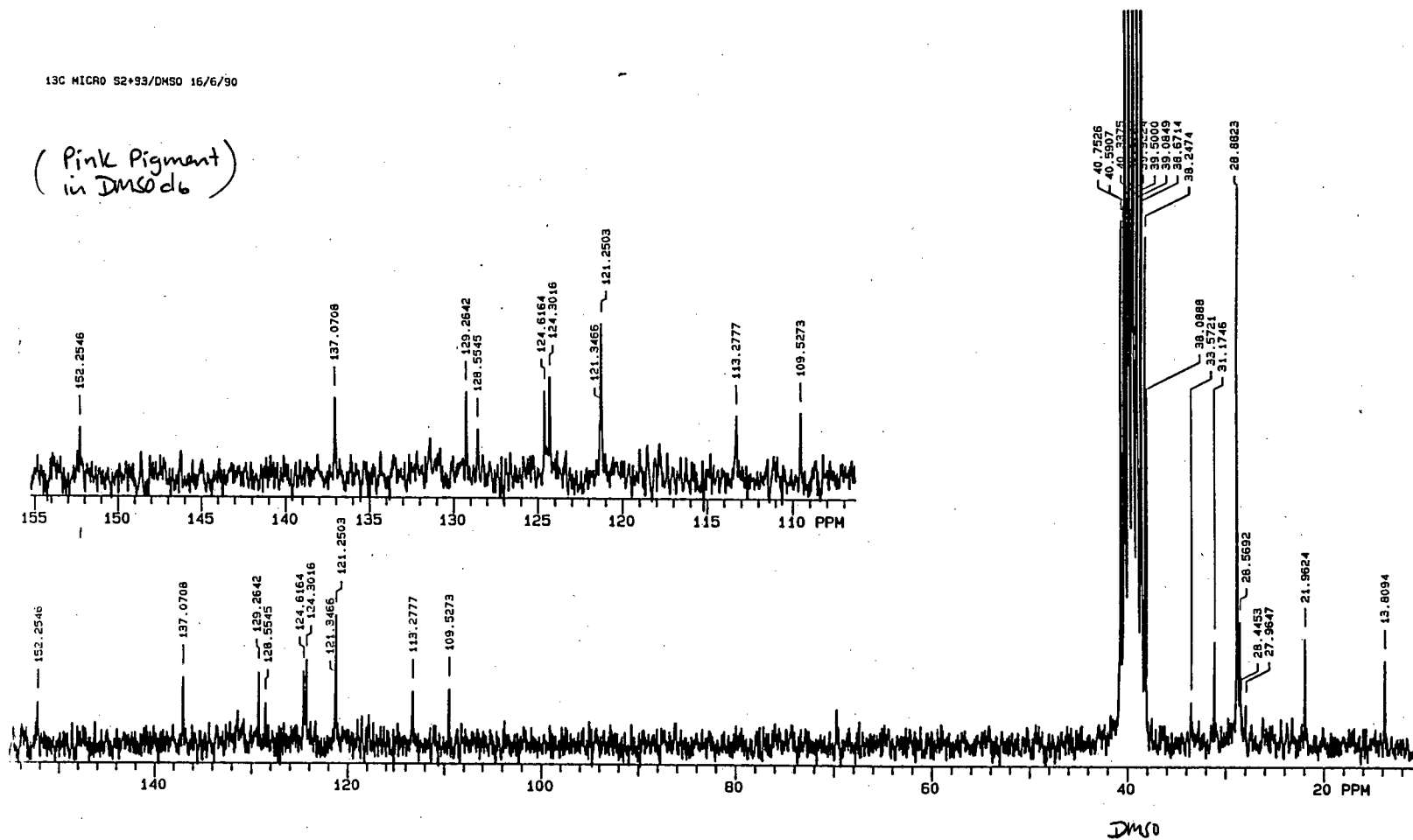
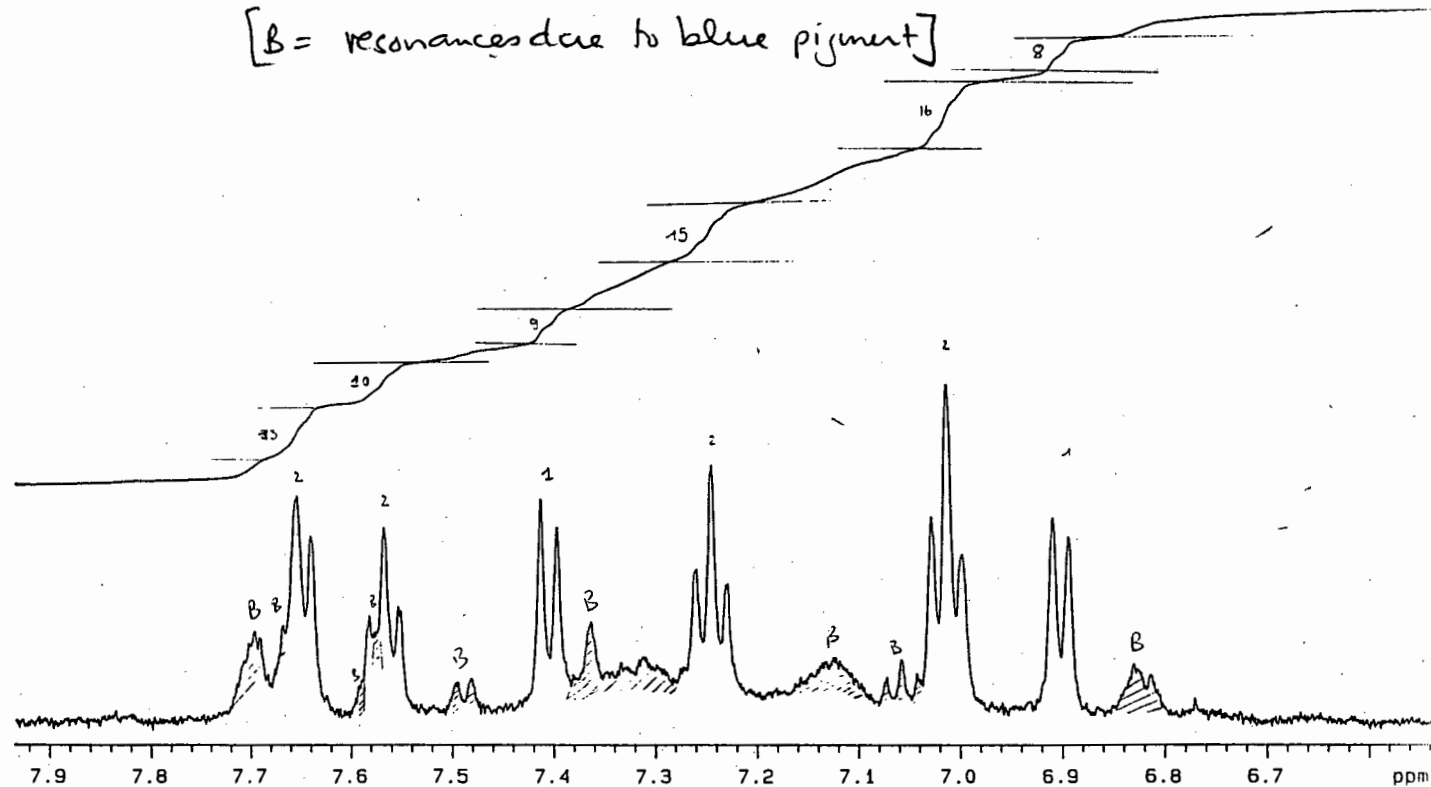


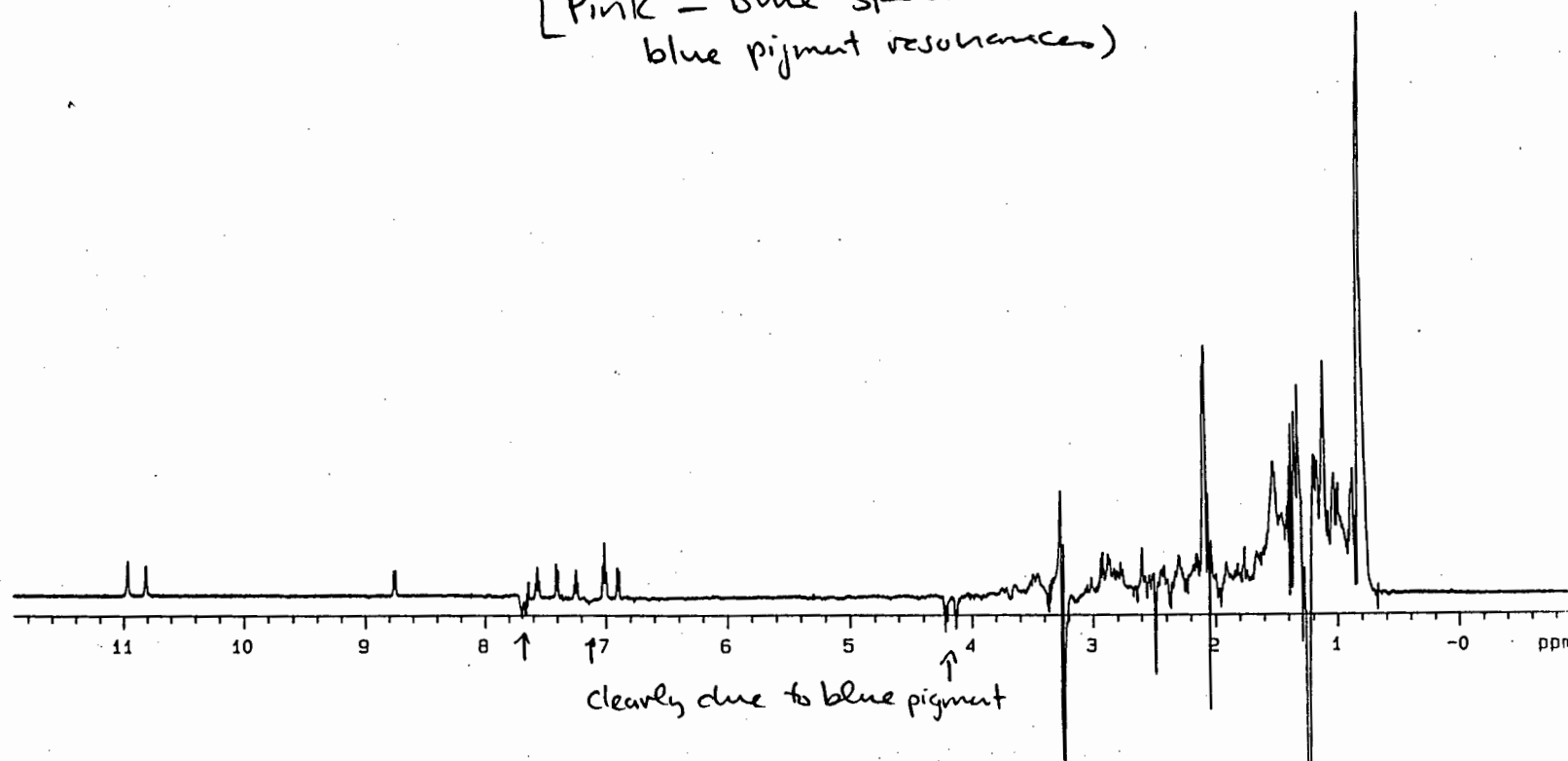
Figure 5(a): Expanded  $^1\text{H}$  spectrum of pink region (500 MHz)

[B = resonances due to blue pigment]



|         |             |              |            |            |          |                  |              |            |            |                 |           |               |       |                |         |         |            |                |           |        |        |     |
|---------|-------------|--------------|------------|------------|----------|------------------|--------------|------------|------------|-----------------|-----------|---------------|-------|----------------|---------|---------|------------|----------------|-----------|--------|--------|-----|
| OBSERVE | Nucleus     | $^1\text{H}$ | Freq.      | 499.84 MHz | DECOUPLE | Nucleus          | $^1\text{H}$ | Offset     | -873.20 Hz | PLOT/PROCESSING | PH        | 65536.00      | RE    | sec            | CO      | sec     | EXPERIMENT | Pulse Sequence | p2p1      | SAMPLE | Number |     |
|         | Spec. Width | 8000.00 Hz   | Offset     | 0 Hz       |          | Mode             | ynn          | Power      | not used   |                 | LB        | not used      | Hz    | AF             | sec     | CCD     |            | Tube O.D.      |           |        | File   | exp |
|         | Acq. Time   | 2.05 sec     | Delay      | 2.00 sec   |          | Modulation: Mode | c            | Freq.      | 200.00 Hz  |                 | Width     | 699.97 Hz/ppm | Start | 3265.58 Hz/ppm | Temp.   | 40.00 C |            | Date           | Aug 20 90 |        |        |     |
|         | Pulse Width | 12.90 sec    | Transients | 32.00      |          | Pulse Width      |              | Power Mode | not used   |                 | Reference |               |       |                | Solvent | DMSO    |            | XL             | 500       |        |        |     |
|         |             |              |            |            |          |                  |              |            |            |                 |           |               |       |                |         |         |            |                |           |        |        | V   |

Figure 5(b) 500 MHz Difference spectrum  
 [Pink - Blue spectra i.e. remove  
 blue pigment resonances]



|            |             |            |            |            |          |                  |      |            |             |                 |           |          |        |       |         |        |         |            |                |       |           |                           |  |
|------------|-------------|------------|------------|------------|----------|------------------|------|------------|-------------|-----------------|-----------|----------|--------|-------|---------|--------|---------|------------|----------------|-------|-----------|---------------------------|--|
| INSTRUMENT | Nucleus     | H1         | Freq.      | 499.84 MHz | DECOUPLE | Nucleus          | H1   | Offset     | -873.20 Hz  | PLOT/PROCESSING | FN        | 65536.00 | K      | RE    | sec     | CO     | sec     | EXPERIMENT | Pulse Sequence | p2pu1 | SAMPLE    | Number                    |  |
|            | Spec. Width | 8000.00 Hz | Offset     | 0 Hz       |          | Mode             | ynn  | Power      | not used db |                 | LB        | not used | Hz     | AF    | sec     | CCO    |         |            | Tube O.D.      | mm    |           | pink_pigment_1h_hires.fid |  |
|            | Acq. Time   | 2.05 sec   | Delay      | 2.00 sec   |          | Modulation: Mode | C    | Freq.      | 200.00 Hz   |                 | Width     | 6497.78  | Hz/ppm | Start | -499.94 | Hz/ppm | Temp.   | 40.00 C    | pink - blue    | Date  | Aug 20 90 |                           |  |
|            | Pulse Width | 12.90 µsec | Transients | 32.00      |          | Pulse Width      | µsec | Power Mode | not used    |                 | Reference |          |        |       |         |        | Solvent | DMSO       | XL             | 500   |           |                           |  |

## APPENDIX D

## LIST OF VECTORS

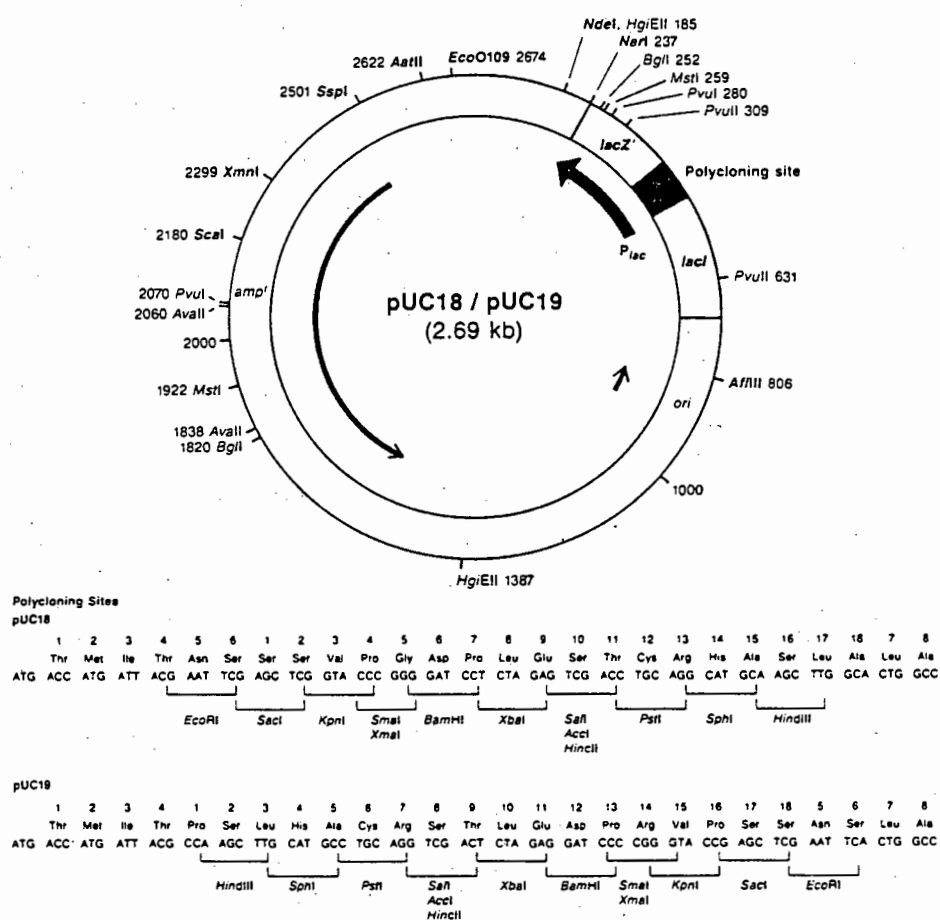
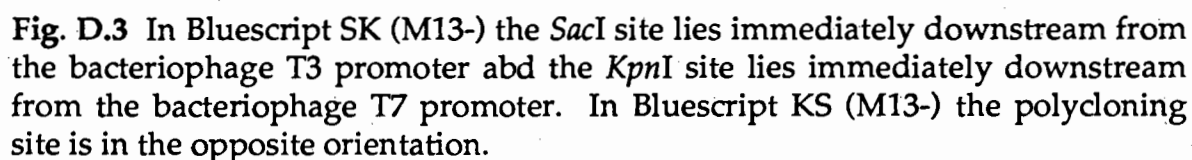
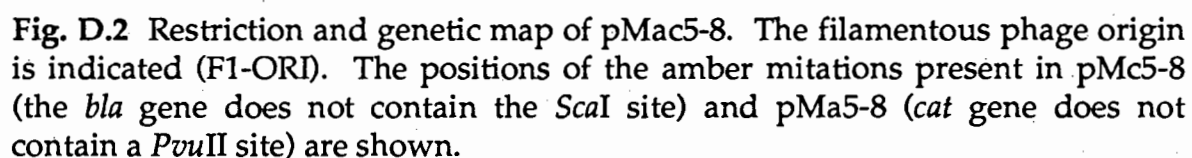


Fig D.1 In pUC18 the *EcoRI* gene lies immediately downstream from *lacOP* and in pUC19 the *HindIII* lies immediately downstream of *lacOP*.



## LITERATURE CITED

- Adams A.N., Bradley S.G.B. 1963 Recombination events in the bacterial genus *Nocardia*. *Science* 140:1392-1394
- Adams J.N. 1963 Nuclear morphogenesis during the developmental cycles of some members of the genus *Nocardia*. *J.Gen.Microbiol.*33:429-435
- Adams J.N. 1964 Recombination between *Nocardia erythropolis* and *Nocardia canicruria*. *J.Bacteriol.*88:865-876
- Adams J.N. 1974 Linkage map and list of markers of *Nocardia erythropolis*. In: *Handbook of Microbiology*. A.I.Laskin and H.A.Lecchevalier (eds). pp. 661-664. CRC Press, Cleveland, Ohio.
- Aiba H., Adhya S., De Crombrughe B. 1981 Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J.Biol.Chem.*256:11905-11910
- Altenbuchner J. 1988 A new *E.coli* cloning vector containing a melanin marker for insertion screening. *Nucleic Acids Res.* 16:8710-
- Apajalahti J.H.A., Salkinoja-Salonen M.S. 1986 Degradation of polychlorinated phenols by *Rhodococcus chlorophenolicus*. *Appl.Microbiol.Biotechnol.*25:62-67
- Apajalahti J.H.A., Salkinoja-Salonen M.S. 1987 Dechlorination and para-hydroxylation of polychlorinated phenols by *Rhodococcus chlorophenolicus*. *J.Bacteriol.*169:675-681
- Asano Y., Fujishiro K., Tani Y., Yamada H. 1982 Aliphatic nitrile hydratase from *Arthrobacter* sp. J-1: purification and characterization. *Agric.Biol.Chem.*46:1165-1174
- Bartel P.L., Zhu C-B., Lampel J.S., Dosch D.C., Connors N.C., Strohl W.R., Beale J.M., Floss H.G. 1990 Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in *Streptomyces*: clarification of actinorhodin gene functions. *J.Bacteriol.*172:4816-4826
- Bell A.A., Wheller M.H. 1986 Biosynthesis and functions of fungal melanins. *Ann.Rev.Phytopathol.*24:411-451
- Bernan V., Filpula D., Herber W., Bibb M., Katz D. 1985 The nucleotide sequence of the tyrosinase gene from *Streptomyces antibioticus* and characterization of the gene product. *Gene* 37:101-110



**Bibb M.J., Cohen S.N.** 1982 Gene expression in *Streptomyces*: construction and application of promoter-probe plasmid vectors in *Streptomyces lividans*. *Mol.Gen.Genet.*187:265-277

**Bibb M.J., Findlay P.R., Johnson M.W.** 1984 The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene* 30:157-166

**Bibb M.J., Ward J.M., Cohen S.N.** 1985 Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to *Streptomyces*. *Mol.Gen.Genet.*199:26-36

**Bilezikian J.P., Kaempfer O.R., Magasanik B.** 1967 Mechanism of tryptophanase induction in *Escherichia coli*. *J.Mol.Biol.*27:495-506

**Billing E., Baker L.A.E.** 1963 Characteristics of *Erwinia*-like organisms found in plant material. *J.Appl.Bacteriol.*26:58-65

**Bolivar F., Rodriguez R.L., Betlach M.C., Boyer H.W.** 1977a Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. *Gene* 2:75-93

**Bolivar F., Rodriguez R.L., Greene P.J., Betlach M.C., Heynecker H.L., Boyer H.W., Crosa J.H., Falkow S.** 1977b Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2:95-113

**Bossi L.** 1983 Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J.Mol.Biol.*164:73-87

**Bossi L., Roth J.R.** 1980 The influence of codon context on genetic code translation. *Nature* 286:123-127

**Botsford J.L., Demoss R.D.** 1971 Catabolite repression of tryptophanase in *Escherichia coli*. *J.Bacteriol.*105:303-312

**Botterman J.** 1986 Expression vector systems and fermentation technology for the production of specific proteins in *Escherichia coli*. PhD thesis, State University of Ghent, Belgium

**Brendel V., Trifonov E.N.** 1984 A computer algorithm for testing potential prokaryotic terminators. *Nucleic Acids Res.* 12:4411-4427

**Brownell G.H.** 1976 Chromosomal location of the  $\phi$ C-prophage in *Nocardia erythropolis*. *J.Gen.Microbiol.*94:217-221

**Brownell G.H.** 1978 Genetic interactions in the genus *Nocardia*. In: Genetics of the Actinomycetales. E.Freersen I.Tarnok (eds). pp. 137-148. JH Thumin-Gustav Fischer Verlag, Stuttgart.

- Brownell G.H.** 1981 A protoplasting and transfecting procedure for *Nocardia*(*Rhodococcus*) species. *Actinomycetes* 16:54-56
- Brownell G.H., Adams J.N.** 1976 Inheritance of nocardiphage  $\phi$ EC in matings of nocardial lysogens. *J.Bacteriol.*126:1104-1107
- Brownell G.H., Adams J.N., Bradley S.G.** 1967 Growth and characterization of nocardiphages for *Nocardia canicruria* and *Nocardia erythropolis* mating types. *J.Gen Microbiol.*47:247-256
- Brownell G.H., Denniston K.** 1984 Genetics of the nocardioform bacteria. In: The biology of the actinomycetes. M.Goodfellow,M.Mordarski,S.T.Williams (eds). pp.201-228. Academic Press, London.
- Brownell G.H., Enquist L.W., Denniston-Thompson K.** 1980 An analysis of the genome of actinophage  $\phi$ EC. *Gene* 12:311-314
- Brownell G.H., Kelly K.L.** 1969 Inheritance of nocardial mating factors. *J.Bacteriol.*99:25-36
- Brownell G.H., Saba J.A., Denniston K., Enquist L.W.** 1982 The development of a *Rhodococcus* actinophage gene cloning system. *Dev.Ind.Microbiol.*24:287-298
- Brown O.R., Clark J.B.** 1966 Pigments of *Nocardia corallina*. *J.Bacteriol.*92:1844-1845
- Burchard R.P., Dworkin M.** 1966.Light-induced lysis and carotenogenesis in *Myxococcus xanthus*. *J.Bacteriol.*91:535-545
- Burchard R.P., Gordon S.A., Dworkin M.** 1966 Action spectrum for the photolysis of *Myxococcus xanthus*. *J.Bacteriol.*91:896-897
- Buttner M.J., Brown N.L.** 1987 Two promoters from the *Streptomyces* plasmid pIJ101 and their expression in *Escherichia coli*. *Gene* 51:179-186
- Casadaban M.J., Martinez-Arias A., Shapira S.K., Chou,J.** 1983  $\beta$ -galactosidase-gene fusions for analyzing gene expression in *Escherichia coli* and yeast. *Methods in Enzymol.*100:293-308
- Chacko C.I., Lockwood J.I., Zabik M.** 1966 Chlorinated hydrocarbon pesticides degradation by microbes. *Science* 154:893-895
- Chatterjee A.K., Gibbins L.N.** 1971 Induction of nonpigmented variants of *Erwinia herbicola* by incubation at supraoptimal temperatures. *J.Bacteriol.*105:107-112
- Chung C.T., Miller R.H.** 1988 A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Res.*16:3580
- Clark-Curtiss J.E., Jacobs W.R., Docherty M.A., Ritchie L.R., Curtiss III R.** 1985 Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. *J.Bacteriol.*161:1093-1102

- Cole S.P., Rudd B.A.M., Hopwood D.A., Chang C., Floss H.G. 1987 Biosynthesis of the antibiotic actinorhodin. *J. Antibiot.* 40:340-347
- Cook A.M., Hutter R. 1986 Ring dechlorination of diethylsimazine by hydrolases from *Rhodococcus corallinus*. *FEMS Microbiol. Letts* 34:335-338
- Cooney J.J., Marks H.W., Smith A.M. 1966 Isolation and identification of canthaxanthin from *Micrococcus roseus*. *J. Bacteriol.* 92:342-345
- Crawford R., Robinson L., Chet M. 1981 <sup>14</sup>C labelled lignins as substrates for the study of lignin biodegradation. In: *Lignin biodegradation, Volume I*. T.K.Kirk, T.Miguchi, M.Chang (eds). pp.61-76. CRC Press, Boca Raton.
- Crockett J.K., Brownell G.H. 1972 Isolation and characterization of a lysogenic strain of *Nocardia erythropolis*. *J. Virol.* 10:737-745
- Cross T. 1982 Actinomycetes: a continuing source of new metabolites. *Dev. Ind. Microbiol.* 23:1-18
- Dabbs E.R. 1987 Fusidic acid resistance in *Rhodococcus erythropolis* due to an inducible extracellular inactivating enzyme. *FEMS Microbiol. Letts* 40:135-138
- Dabbs E.R., Sole G.J. 1988 Plasmid-borne resistance to arsenate, arsenite, cadmium and chloramphenicol in a *Rhodococcus* species. *Mol. Gen. Genet.* 211:148-154
- D'Aoust J.Y., Gerber N.N. 1974 Isolation and purification of prodigiosin from *Vibrio psychroerythreus*. *J. Bacteriol.* 118:756-757
- Dauenhauer S.A., Hull R.A., Williams R.P. 1984 Cloning and expression in *Escherichia coli* of *Serratia marcescens* genes encoding prodigiosin biosynthesis. *J. Bacteriol.* 158:1128-1132
- Davis P.J., Rizzo J.D. 1982 Microbial transformations of Warfarin: stereoselective reduction by *Nocardia corallina* and *Arthrobacter* species. *Appl. Environ. Microbiol.* 43:884-890
- Della-Cioppa G., Garger S.J., Sverlow G.G., Turpen T.H., Grill L.K. 1990 Melanin production in *Escherichia coli* from a cloned tyrosinase gene. *Bio/Technology* 8:634-638
- Deng Z., Kieser T., Hopwood D.A. 1986 Expression of a *Streptomyces* plasmid promoter in *Escherichia coli*. *Gene* 43:295-300
- Desomer J., Dhaese P., Van Montagu M. 1987 Isolation of 237  $\phi$ CF1, a temperate bacteriophage from *Corynebacterium fascians*. *Arch. Int. Physiol. Biochim.* 95:B21
- Desomer J., Dhaese P., Van Montagu M. 1988 Conjugative transfer of cadmium resistance plasmids in *Rhodococcus fascians* strains. *J. Bacteriol.* 170:2401-2405

- Desomer J., Dhaese P., Van Montagu M. 1990 Transformation of *Rhodococcus fascians* by high-voltage electroporation and development of *R. fascians* cloning vectors. *Appl. Environ. Microbiol.* 56:2818-2825
- Devereux J., Haerberli P., Smithies O. 1984 A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-395
- Dreyfuss M. 1988 What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs. *J. Mol. Biol.* 204:79-94
- Eaton R.W., Timmis K.N. 1986 Characterization of a plasmid-specified pathway for catabolism of isopropyl benzene in *Pseudomonas putida* RE204. *J. Bacteriol.* 168:123-131
- Enquist L.W., Bradley S.G. 1971 Characterization of deoxyribonucleic acid from *Streptomyces venezuelae* spores. *Dev. Ind. Microbiol.* 12:225-236
- Ensley B.D., Gibson D.T., Laborde A.L. 1982 Oxidation of naphthalene by a multicomponent enzyme system from *Pseudomonas* sp. strain NCIB 9816. *J. Bacteriol.* 149:948-954
- Ensley B.D., Ratzkin B.R., Osslund T.D., Simon M.J., Wackett L.P., Gibson D.T. 1983 Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* 222:167-169
- Ernst J.F. 1988 Codon usage and gene expression. *Tibtech* 6:196-199
- Feitelson J.S., Hopwood D.A. 1983 Cloning of a *Streptomyces* gene for an *o*-methyl transferase involved in antibiotic biosynthesis. *Mol. Gen. Genet.* 190:394-398
- Feitelson J.S., Malpartida F., Hopwood D.A. 1985 Genetic and biochemical characterisation of the red gene cluster of *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* 131:2431-2441
- Feitelson J.S., Sinha A.M., Coco E.A. 1986 Molecular genetics of red biosynthesis in *Streptomyces*. *J. Nat. Prod.* 49:988-994
- Feng J.S., Webb J.W., Tsang J.C. 1982 Enhancement by sodium dodecyl sulphate of pigment formation in *Serratia marcescens* 08. *Appl. Environ. Microbiol.* 43:850-853
- Ferguson J.A., Korte F. 1981 Transformation of aldrin by soil microorganisms. *Appl. Environ. Microbiol.* 34:7-15
- Floss H.G., Cole S.P., He X-G., Rudd B.A.M., Duncan J., Fujii I., C-J., Keller P.J. 1985 Biosynthesis of polyketide antibiotics. H. Kleinkauf, H. von Dohren, H. Dornauer, G. Nesemann (eds). pp.283-304. VCH Verlagsgesellschaft mbH, Weinheim, FRG.
- Freundlich M., Lichstein H.C. 1960 Inhibitory effect of glucose on tryptophanase. *J. Bacteriol.* 80:633-640

- Fujimori E., Livingston R. 1957 Interaction of chlorophyll in its triplet state with oxygen, carotene, etc. *Nature (London)* 180:1036-1038
- Furuhashi K., Shintani M., Takagi M. 1986 Effects of solvents on the production of epoxides by *Nocardia corallina* B-276. *Appl. Microbiol. Biotechnol.* 23:218-223
- Furuhashi K., Taoka A., Uchida S., Karube I., Suzuki S. 1981 Production of 1,2,-epoxyalkanes from 1-alkanes by *Nocardia corallina* B-276. *Eur. J. Appl. Microbiol. Biotechnol.* 12:39-45
- Gandhi N.M., Nazareth J., Divekar P.V., Kohl H., De Souza N.J. 1973 Magnesidin, a novel magnesium-containing antibiotic. *J. Antibiot.* 26:797-798
- Gantotti B.V., Beer S.V. 1982 Plasmid-borne determinants of pigmentation and thiamine prototrophy in *Erwinia herbicola*. *J. Bacteriol.* 151:1627-1629
- Gibson D.T., Hensley M., Yoshioka H., Mabry T.J. 1970 Formation of (+)-*cis*-2,3-dihydroxy-1-methylcyclohexa-4,6-diene from toluene by *Pseudomonas putida*. *Biochem.* 9:1626-1630
- Gibson D.T., Koch J.R., Kallio R.E. 1968 Oxidative degradation of aromatic hydrocarbons by microorganisms. I. Enzymatic formation of catechol from benzene. *Biochem.* 7:2653-2662
- Gold L., Pribnow D., Schneider T., Shinedling S., Singer B.S., Stromo G. 1981 Translational initiation in prokaryotes. *Annu. Rev. Microbiol.* 35:365-403
- Goodfellow M., Alderson G. 1977 The actinomycete genus *Rhodococcus*: a home for the "rhodochrous" complex. *J. Gen. Microbiol.* 100:99-122
- Goodfellow M., Minnikin D.E. 1977 Nocardiform bacteria. *Annu. Rev. Microbiol.* 31:159-180
- Goodman R.N. 1965 *In vitro* and *in vivo* interactions between components of mixed bacterial cultures isolated from apple buds. *Phytopathol.* 55:217-221
- Gouy M. 1987 Codon contexts in enterobacterial and coliphage genes. *Mol. Biol. Evol.* 4:426-444
- Gray P.H.H. 1927 The formation of indigotin from indol by soil bacteria. *Proc. R. Soc. Lond. B* 102:263-279
- Green J.A., Rappoport D.A., Williams R.P. 1956 Studies on pigmentation of *Serratia marcescens*. II. Characterization of the blue and the combined red pigments of prodigiosin. *J. Bacteriol.* 72:483-487
- Griffiths M., Sistrom M.R., Cohen-Bazire G. 1955 Function of carotenoids in photosynthesis. *Nature (London)* 176:1211-1214

- Häggeblom M., Apajalahti J., Salkinoja-Salonen M. 1986 Metabolism of chloroguaiacols by *Rhodococcus chlorophenolicus*. Appl.Microbiol.Biotechnol.24:397-404
- Haider K., Jugnow G., Kohnen R., Lim S.V. 1981 Degradation of chlorinated benzenes, phenols and cyclohexane derivatives by soil bacteria under aerobic conditions. In: Decomposition of toxic and nontoxic organic compounds in soil. M.R.Overcash (ed). pp.207-226. Ann Arbor Science, Ann Arbor.
- Hart S., Kirby R., Woods D.R. 1990 Structure of a *Rhodococcus* gene encoding pigment production in *Escherichia coli*. J.Gen.Microbiol.136:1357-1363
- Henikoff S. 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359
- Hill R.T., Hart S.L., Illing N., Kirby R., Woods D.R. 1989 Cloning and expression of *Rhodococcus* pigment genes in *Escherichia coli*. J.Gen.Microbiol.135:1507-1513
- Hirsch P., Alexander M. 1960 Microbial decomposition of halogenated propionic and acetic acid. Can.J.Microbiol.6:241-249
- Hobbs G., Frazer C.M., Gardner D.C.J., Flett F., Oliver S.G. 1990 Pigmented antibiotic production by *Streptomyces coelicolor* A3(2): kinetics and the influence of nutrients. J.Gen.Microbiol.136:2291-2296
- Honda Y., Sakai H., Komano T. 1988. Two single stranded DNA initiation signals located in the *oriV* region of plasmid RSF1010. Gene 68:221-228
- Horinouchi S., Beppu T. 1985 Construction and application of a promoter-probe plasmid that allows chromogenic identification in *Streptomyces lividans*. J.Bacteriol.162:406-412
- Hummel W., Schutte H., Schmidt E., Wandrey C., Kula M-R. 1987 Isolation of L-phenylalanine dehydrogenase from *Rhodococcus* sp. M4 and its application for the production of L-phenylalanine. Appl.Microbiol.Biotechnol.26:409-416
- Ikehata O., Nishiyama M., Horinouchi S., Beppu T. 1989 Primary structure of nitrile hydratase deduced from the nucleotide sequence of a *Rhodococcus* species and its expression in *Escherichia coli*. E.J.Biochem.181:563-570
- Ikemura T. 1985 Codon usage and tRNA content in unicellular and multicellular organisms. Mol.Biol.Evol.2:13-34
- Ikuta N., Souza M.B.N., Valencia F.F., Castro M.E B., Schenberg A.C.G., Pizzirani-Kleiner A., Astolfi-Filho S. 1990 The  $\alpha$ -amylase gene as a marker for gene cloning: direct screening of recombinant clones. Bio/Technology.8:241-242
- Irie S., Doi S., Yorifuji T., Takagi M., Yano K. 1987 Nucleotide sequencing and characterization of the genes encoding benzene oxidation enzymes of *Pseudomonas putida*. J.Bacteriol.169:5174-5179

- Irie S., Shirai K., Doi S., Yorifuji T. 1987b Cloning of genes encoding oxidation of benzene in *Pseudomonas putida* and their expression in *Escherichia coli* and *P. putida*. Agric.Biol.Chem.51:1489-1493
- Ish-Horowicz D., Burke J.F. 1981 Rapid and efficient cosmid vector cloning. Nucleic Acids Res.9:2989-2998
- Jaeger E., Eggeling L., Sahm H. 1981 Partial purification and characterization of coniferyl alcohol dehydrogenase from *Rhodococcus erythropolis*. Curr.Microbiol.6:333-336
- Jaurin B., Cohen S.N. 1985 *Streptomyces* contain *Escherichia coli* type A+T-rich promoters having novel structural features. Gene 39:191-201
- Jefferson R.A., Burgess S.M., Hirsh D. 1986  $\beta$ -glucuronidase from *Escherichia coli* as a gene-fusion marker. Proc.Natl.Acad.Sci.USA 83:8447-8451
- Katz E., Thompson C.J., Hopwood D.A. 1983 Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. J.Gen.Microbiol.129:2703-2714
- Kieser T., Moss M.T., Dale J.W., Hopwood D.A. 1986 Cloning and expression of *Mycobacterium bovis* BCG DNA in *Streptomyces lividans*. J.Bacteriol.168:72-80
- Kirby R., Usdin K. 1985 The isolation and restriction mapping of a mini-plasmid from the actinomycete *Nocardia corallina*. FEMS Microbiol.Letts 27:57-60
- Kobayashi M., Yanaka N., Nagasawa T., Yamada H. 1990 Purification and characterization of a novel nitrilase of *Rhodococcus rhodochrous* K22 that acts on aliphatic nitriles. J.Bacteriol.172:4807-4815
- Kozak M. 1983 Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. Microbiol.Revs.47:1-45
- Kramer W., Drutsa V., Jansen H-W., Kramer B., Pflugfelder M., Fritz H-J. 1984 The gapped duplex DNA approach to oligonucleotide-directed mutation construction. Nucleic Acids Res.12:9441-9456
- Kurane R., Suzuki T., Fukuoka S. 1984 Purification and some properties of a phthalate ester hydrolyzing enzyme from *Nocardia erythropolis*. Appl.Microbiol.Biotechnol.20:378-383
- Kurane R., Suzuki T., Takahara Y. 1979 Removal of phthalate esters by activated sludge inoculated with a strain of *Nocardia erythropolis*. Agric.Biol.Chem.43:421-427
- Kurane R., Suzuki T., Takahara Y. 1980 Metabolic pathway of phthalate esters by *Nocardia erythropolis*. Agric.Biol.Chem.44:523-527

**Kuwuhara M.** 1981 Metabolism of lignin-related compounds by bacteria. In: Lignin biodegradation, Vol II. T.K.Kirk, T.Higuchi, H.Chang (eds). pp.127-146. CRC Press, Boca Raton.

**Labidi A., David H.L., Roulland-Dussoix D.** 1985 Cloning and expression of mycobacterial plasmid DNA in *Escherichia coli*. FEMS Microbiol.Letts 30:221-225

**Laemmli U.K.** 1970 Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685

**Lawson E.N., Gantotti B.V., Starr M.P.** 1982 A 78-megadalton plasmid occurs in avirulent strains as well as virulent strains of *Corynebacterium fascians*. Curr.Microbiol.7:327-332

**Lee Y-H.W., Chen B.F., Wu S.Y., Leu W.M., Lin J.J., Chen C.W., Lo S.J.** 1988 A trans-acting gene is required for the phenotype expression of a tyrosinase gene in *Streptomyces*. Gene 65:71-81

**Leinfelder W., Zehelein E., Mandrand-Berthelot M-A., Böck A.** 1988 Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature 331:723-725

**Loper J.E., Orser C.S., Panapoulos N.J., Schroth M.N.** 1984 Genetic analysis of fluorescent pigment production in *Pseudomonas syringae* pv. *syringae*. J.Gen.Microbiol.130:1507-1515

**Malpartida F., Hopwood D.A.** 1984 Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature 309:462-464

**Malpartida F., Hopwood D.A.** 1986 Physical and genetic characterization of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). Mol.Gen.Genet.205:66-73

**Maniatis T., Fritsch E.F., Sambrook J.** 1982 Molecular cloning: a laboratory manual. Cold Spring Harbor, New York, Cold Spring Harbour Laboratory.

**Mathews M.M.** 1963 Comparative study of lethal photosensitization of *Sarcina lutea* by 8-methoxypsoralen and by toluidine. J.Bacteriol.85:322-328

**Mathews M.M., Sistrom W.R.** 1959 Function of carotenoid pigments in non-photosynthetic bacteria. Nature (London) 184:1892-1893

**Mikami Y., Takahashi K., Fukushima K., Yazawa K., Arai T., Kubo A., Saito N., Kawakami N.** 1987 A new pigment from *Streptomyces lavendulae*. J.Antibiot.60:385-387

**Miller H.J.** 1980 Recent observations on leafy gall in Lilliacae and some other families. Neth.J.Plant Pathol.86:55-68



- Miller J.H., Albertini A.M. 1983 Effects of surrounding sequence on the suppression of nonsense codons. *J.Mol.Biol.*164:59-71
- Morgan W.D., Bear D.G., Litchman B.L., von Hippel P.H. 1985 RNA sequence and secondary structure requirements for rho-dependant transcription termination. *Nucleic Acids Res.* 13:3739-3754
- Murai N. 1981 Cytokinin biosynthesis and its relationship to the presence of plasmids in strains of *Corynebacterium fascians*. In: Metabolism and molecular activities of cytokinins. J.Guern and C.Peaud-Lenoed (eds). Springer-Verlag KG, Berlin.
- Murai N., Skoog F., Doyle M.E., Hanson R.S. 1980 Relationship between cytokinin production, presence of plasmids and fasciation caused by strains of *Corynebacterium fascians*. *Proc.Natl.Acad.Sci.USA* 77:619-623
- Murdock D., Sendar C., Ensley B. 1990 Channeling substrate carbon to specialty chemicals: synthesis of indigo from glucose. 6th International Symposium on Genetics of Industrial Microorganisms, Strasbourg, 1990, p.37.
- Muto A., Osawa S. 1987 The guanine and cytosine of genomic DNA and bacterial evolution. *Proc.Natl.Acad.Sci.USA* 84:166-169
- Nakayama M., Fujita N., Ohama T., Osawa S., Ishihama A. 1989 *Micrococcus luteus*, a bacterium with a high genomic G+C content, contains *Escherichia coli*-type promoters. *Mol.Gen. Genet.*218:384-389
- Olsson O., Koncz C., Szalay A.A. 1988 The use of the *lux A* gene of the bacterial luciferase operon as a reporter gene. *Mol.Gen.Genet.*215:1-9
- Oshima T.O., Kawai S., Egami F. 1965 Oxidation of indole to indigotin by *Pseudomonas indoloxidans*. *J.Biochem.*58:259-263
- Pardee A.B., Jacob F., Monod J. 1959 The genetic control and cytoplasmic expression of inducibility in the synthesis of  $\beta$ -galactosidase by *Escherichia coli*. *J.Mol.Biol.*1:165-178
- Park Y-H., Hori H., Suzuki K-I., Osawa S., Komayata K. 1988 Nucleotide sequence of 5S ribosomal RNA from *Rhodococcus erythropolis*. *Nucleic Acids Res.*15:365
- Peczynska-Czoch W., Mordarski M. 1984 Transformation of xenobiotics. In: The biology of the actinomycetes. M.Goodfellow, M.Mordarski, S.T.Williams (eds). pp.287-336. Academic Press, London.
- Perry K.L., Simonitch T.A., Harrison-Lavoie K.J., Liu S-T.1986 Cloning and regulation of *Erwinia herbicola* pigment genes. *J.Bacteriol.*168:607-612
- Petersen G.B., Stockwell P.A., Hill D.F. 1988 Messenger RNA recognition in *Escherichia coli*: a possible second site of interaction with 16S ribosomal RNA. *EMBO J.*7:3957-3962

**Rast H.G., Engelhardt G., Diegler W., Wallhoffer P.R.** 1980 Bacterial degradation of model compounds for lignin and chlorophenol derived lignin bound residues. *FEMS Microbiol.Letts* 8:259-263

**Raymond R.L.** 1971 Hydroxy-phenyl- $\alpha$ -ketobutyric acids. United States Patent Office. Patent no. 3,392,846.

**Raymond R.L., Jameson V.W.** 1971 Biochemical activities of *Nocardia*. *Adv.Appl.Microbiol.*14:93-122

**Reh M., Schlegel H.G.** 1981 Hydrogen autotrophy as a transferable genetic character of *Nocardia opaca* 1b. *J.Gen. Microbiol.*126:327-336

**Riverolezcano O., Anguitacastillo J., Lopeznieto J., Naharrocarasco G.** 1990 Cloning and expression in *Escherichia coli* of tryptophan genes from *Streptomyces griseus* IMRU-3570. *FEMS Microbiol.Letts* 68:201-205

**Rosenberg M., Court D.** 1979 Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu.Rev.Genet.*13:319-353

**Rudd B.A.M., Hopwood D.A.** 1979 Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J.Gen. Microbiol.*119:333-340

**Rudd B.A.M., Hopwood D.A.** 1980 A pigmented mycelial antibiotic in *Streptomyces coelicolor*:control by a chromosomal gene cluster. *J.Gen.Microbiol.*119:333-340

**Sallis P.J., Armfield S.J., Bull A.T., Hardman D.J.** 1990 Isolation and characterization of a haloalkane halidohydrolase from *Rhodococcus erythropolis* Y2. *J.Gen. Microbiol.*136:115-120

**Salser W.** 1977 Globin mRNA sequences: analysis of base pairing and evolutionary implications. *Cold Spring Harbor Symposia on Quantitative Biology* 42:985-1002

**Sandmann G., Woods W.S., Tuveson R.W.** 1990 Identification of carotenoids in *Erwinia herbicola* and in a transformed *Escherichia coli* strain. *FEMS Microbiol.Letts* 71:77-82

**Sanger F., Nicklen S., Coulson A.R.** 1977 DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.USA* 74:5463-5467

**Sebek O.K., Jager H.** 1962 Divergent pathways of indole metabolism in *Chromobacterium violaceum*. *Nature* 24:793-795

**Sensfuss C., Reh M., Schlegel H.G.** 1986 No correlation exists between the conjugative transfer of the autotrophic character and that of plasmids in *Nocardia opaca* strains. *J.Gen.Microbiol.*132:997-1007

- Setchell C.H., Bonner J.F., Wright S.J., Caunt P., Baker P.B.** 1985 Microbial transformation of squalene: formation of a novel ketone from squalene by a *Rhodococcus* sp. Appl.Microbiol.Biotechnol.21:255-257
- Shewan J.M., McMeekin T.A.** 1983 Taxonomy (and ecology) of *Flavobacterium* and related genera. Annu. Rev. Microbiol.37:233-252
- Shine J., Dalgarno L.** 1974 The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc.Natl. Acad.Sci.USA 71:1342-1346
- Shpaer E.G.** 1986 Constraints on codon context in *Escherichia coli* genes: their possible role in modulating the efficiency of translation. J.Mol.Biol.188:555-564
- Sobin B., Stahly G.L.** 1942 Chromogenesis by variants of *Staphylococcus aureus*. J.Bacteriol.92:1578-1579
- Stanssens P., Opsomer C., Mckeown Y.M., Kramer W., Zabeau M., Fritz H-J.** 1989 Efficient oligonucleotide-directed construction of mutations in expression vectors by the gapped duplex DNA method using alternating selectable markers. Nucleic Acids Res.17:4441-4454
- Stapp C.** 1981 Bacterial plant pathogens. Oxford University Press, Oxford.
- Stees J.L., Brown O.R., Grigsby R.D.** 1969 Structural features of the major pigment component of *Nocardia corallina*. Microbios 2:199-200
- Stormo G.D.** 1986 Translation initiation. In: Maximizing gene expression. W.Reznikoff, L.Gold (eds). pp.195-224. Butterworth Publishers, Stoneham, Mass.
- Strauss D.G., Berger U., Fleck W.F.** 1983 Pigments from microorganisms. Chemistry and biotechnology of biologically active natural products. 2nd International Conference, Budapest, Hungary.
- Tabor S.** 1987 Phosphatase and kinases. In: Current Protocols in Molecular Biology. F.M.Ausebel, R.Brent, R.E.Kingston, D.D.Moore, J.G.D.Seidman, J.A.Smith, K.Struhl (eds) pp.3.10.1-3.10.2. Greene Publishing Associates and Wiley Interscience, New York.
- Takaichi S., Ishidsu J-I., Seki T., Fukada S.** 1990 Carotenoid pigments from *Rhodococcus rhodochrous* RNMS1: two monocyclic carotenoids, a carotenoid monoglycoside and carotenoid glycoside monoesters. Agric.Biol.Chem.54:1931-1937
- Tan H-M., Mason J.R.** 1990 Cloning and expression of the plasmid-encoded benzene dioxygenase genes from *Pseudomonas putida* ML2. FEMS Microbiol.Letts 72:259-264
- Thole J.E.R., Dauwerse H.G., Das P.K., Groothuis D.G., Schouls L.M., van Embden J.D.A.** 1985 Cloning of *Mycobacterium bovis* BCG DNA and expression of antigens in *Escherichia coli*. Infect.Immun.50:800-806

- Trojanowski J., Haider K., Sundman V. 1977 Decomposition of  $^{14}\text{C}$ -labelled lignin and phenols by a *Nocardia* sp. *Archiv fur Mikrobiologie* 114:149-153
- Tseng H-C., Lin C-K., Hsu B-J., Leu W-M., Wu Lee Y-H., Chiou S-J., Hu N-T., Chen C.W. 1990 The melanin operon of *Streptomyces antibioticus*: expression and use as a marker in Gram-negative bacteria. *Gene* 86:123-128
- Tuveson R.W., Larson R.A., Kagan J. 1988 Role of cloned carotenoid genes expressed in *Escherichia coli* in protecting against inactivation by near-UV light and specific phototoxic molecules. *J.Bacteriol.*170:4675-4680
- Vantomme R., Elia S., Swings J., de Ley J. 1982 *Corynebacterium fascians* (Tilford, 1936) Dowson 1942, the causal agent of leafy gall on lily crops in Belgium. *Parasitica* 38:183-192
- Viera J., Messing J. 1982 The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19:259-268
- Vogt Singer M.E., Finnerty W.R. 1988 Construction of an *Escherichia coli*-*Rhodococcus* shuttle vector and plasmid transformation in *Rhodococcus* sp. *J.Bacteriol.*170:638-645
- Wada K-N., Aota S-I., Suchiya R., Ishibashi F., Gojobori T., Toshimichi I. 1990 Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res.(suppl.)* 18:2367-2411
- Ward D.F., Yudkin M.D. 1976 Mutations in *Escherichia coli* that relieve catabolite repression of tryptophanase synthesis. Tryptophanase promoter-like mutations. *J.Gen.Microbiol.*92:133-137
- Watanabe I., Satoh Y., Enomoto K. 1987a Screening, isolation and taxonomical properties of microorganisms having acrylonitrile hydrating activity. *Agric.Biol.Chem.*51:3193-3199
- Watanabe I., Satoh Y., Enomoto K., Seki S., Sakashita K. 1987b Optimal conditions for cultivation of *Rhodococcus* sp. N-774 and for conversion of acrylonitrile to acrylamide by resting cells. *Agric.Biol.Chem.*51:3201-3206
- Watson J.D., Hopkins N.H., Roberts J.W., Steitz J.A., Weiner A.M. 1987 *Molecular Biology of the Gene*, 4<sup>th</sup> edition, Volume 1. pp.409-411. Benjamin/Cummings, California.
- Westpheling J., Raney M., Losick R. 1985 RNA polymerase heterogeneity in *Streptomyces coelicolor*. *Nature* 313:22-27
- Williams R.P., Green J.A., Rappaport D.A. 1955 Studies on pigmentation of *Serratia marcescens*. I. Spectral and paper chromatographic properties of prodigiosin. *J.Bacteriol.*71:115-120

- Woods D.R., Mossmann T.R., Hanson S., Hendry S. 1971 Pigmentation and acriflavine resistance in *Serratia marcescens*. J.Bacteriol.108:765-769
- Woods D.R., Schauder V.R., Waddington P.B. 1973 Acriflavine uptake and resistance in *Serratia marcescens* cells and spheroplasts. J.Bacteriol.114:59-64
- Woods D.R., Sparrow O.F., Mossmann T.R., Hendry D.A. 1971 Induction of resistance and pigmentation mutations by acriflavine in *Serratia marcescens*. Proc.4th Cong.S.A. Genet.Soc.(Pretoria 1970)
- Yanagimoto M., Matsumoto K., Mori K. 1988 IM2; a new inducer of blue pigment production in *Streptomyces* sp. MAFF 10-06015. J.Ferment.Technol.66:1-6
- Yanisch-Perron C., Vieira J., Messing J. 1985 Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119
- Yarus M., Folley L.S. 1985 Sense codons are found in specific contexts. J.Mol.Biol.182:529-540
- Yeh W-K., Gibson D.T., Liu T-N. 1977 Toluene dioxygenase: a multicomponent enzyme system. Biochem.Biophys.Res.Comm.78:401-410
- Zabeau M., Stanley K. 1982 Enhanced expression of cro- $\beta$ -galactosidase fusion proteins under the control of the  $P_R$  promoter of bacteriophage lambda. EMBO J.1:1217-1224
- Zamanian M., Mason J.R. 1987 Benzene dioxygenase in *Pseudomonas putida*. Subunit composition and immuno-cross-reactivity with other aromatic dioxygenases. Biochem.J.224:611-616
- Ziffer H., Jerina D.M., Gibson D.T., Kobal V.M. 1973 Absolute stereochemistry of the dextro *cis*-1,2-dihydroxy-3-methylcyclohexa-3,5-diene produced from toluene by *Pseudomonas putida*. J.Am.Chem.Soc 95:4048-4049
- Zinoni F., Birkmann A., Leinfelder W., Böck A. 1987 Cotranslational insertion of selenocysteine into formate dehydrogenase of *Escherichia coli* directed by a UGA codon. Proc.Natl.Acad.Sci.USA 84:3156-3160
- Zylstra G.J., Gibson D.T. 1989 Toluene degradation by *Pseudomonas putida* F1. J.Biol.Chem.264:14940-14946
- Zylstra G.J., McCombie W.R., Gibson D.T., Finette B.A. 1988 Toluene degradation by *Pseudomonas putida* F1: genetic organization of the *tod* operon. Appl.Environ. Microbiol.54:1498-1503